Knowledge management and human trafficking: using conceptual knowledge representation, text analytics and open-source data to combat organized crime

BREWSTER, Benjamin <http://orcid.org/0000-0003-3536-5862>, POLOVINA, Simon <http://orcid.org/0000-0003-2961-6207>, RANKIN, Glynn and ANDREWS, Simon <http://orcid.org/0000-0003-2094-7456>

Citation:

BREWSTER, Benjamin, POLOVINA, Simon, RANKIN, Glynn and ANDREWS, Simon (2014). Knowledge management and human trafficking: using conceptual knowledge representation, text analytics and open-source data to combat organized crime. In: HERNANDEZ, Nathalie, JÄSCHKE, Robert and CROITORU, Madalina, (eds.) Graph-Based Representation and Reasoning. Lecture Notes in Computer Science (8577). Springer International Publishing, 104-117. [Book Section]

Knowledge Management and Human Trafficking: Using Conceptual Knowledge Representation, Text Analytics and open-source data to Combat Organized Crime

Ben Brewster[1], Simon Polovina[1], Glynn Rankin[2], and Simon Andrews[1]

[1]CENTRIC, Sheffield Hallam University, UK
{B.Brewster, S. Polovina, S. Andrews}@SHU.ac.uk

[2]Rankin Kinsella Associates, UK
GlynnRankin@RKA.org.uk

Abstract. Globalization, the ubiquity of mobile communications and the rise of the web have all expanded the environment in which organized criminal entities are conducting their illicit activities, and as a result the environment that law enforcement agencies have to police. This paper triangulates the capability of open-source data analytics, ontological knowledge representation and the wider notion of knowledge management (KM) in order to provide an effective, inter-disciplinary means to combat such threats, thus providing law enforcement agencies (LEA's) with a foundation of competitive advantage over human trafficking and other organized crime.

1 Introduction

Globalization, the ubiquity of mobile communications and the emergence of the web have all aided in the stimulation of trade links, international communications and increased global mobility [1]. Despite the positive potential of this new global, digital environment, it also provides organized criminal entities with new avenues to exploit when conducting illicit activities, and as a result provides law enforcement agencies (LEA's) with an ever-widening environment to police. The new environment defined by mobile communications and the web provides LEAs with a valuable open-source platform that can be applied to enhance existing intelligence based investigations. The use of analytical approaches such as natural language processing (NLP), sentiment analysis and other text mining techniques enables LEAs to follow the breadcrumb trail of evidence leading towards illegal activity.

In 2012, social media service Twitter reported that on average more than 500 million posts were made to its service per day, and its popularity is continuing to grow year over year [2]. In addition to the sustained ubiquity of established social media, new services continue to emerge, targeting differing market segments and offering varying forms of functionality. The popularity of these services, along with other web sources presents a potentially unrivalled intelligence platform (in terms of size) that LEAs can access in order to combat and prevent a variety of organized criminal threats. Ontological approaches to conceptual knowledge representation and the wider philosophy of knowledge management (KM) can provide the infrastructure needed for the intelligence-led operations of LEAs to utilize open source-data.

The EU, FP7 funded ePOOLICE project aims to make use of open-source repositories such as the web to identify, prevent and even predict emergent organized crime threats to better inform LEAs and the actions and decisions they make in relation to these threats. Through the application and development of an ontological knowledge-base, the system will make use of semantically managed domain expertise, provided by LEA partners in order assess the validity of any data mined from open-source repositories [3]. To facilitate the mining of open-source data, ePOOLICE will use NLP text mining techniques such as concept extraction and content categorization in order to manage the acquisition, identification and filtration of relevant data crawled online, identifying poignant indicators and information. 'Privacy by Design' is also implemented; most notably by anonymizing or removing potentially sensitive data that could be used to explicitly identify individuals. To enable knowledge to be effectively managed, domain expertise such as indicators and their relationships will be codified using conceptual graphs and embedded within a larger knowledgebase.
The triangulation between open-source data analytics, ontological knowledge representation and the wider notion of KM potentially provides LEAs with a foundation of competitive advantage over a range organized criminal threats such as drug trafficking, cyber-crime, terrorism and human trafficking; An advantage not easily reciprocated by criminal groups [4]**.** In this paper we outline the growing problem of human trafficking, and present an overview of the potential capability of ePOOLICE in utilizing open-source data through the use of KM, conceptual graphs and text analytics in order to aid law enforcement in strategically combatting the human trafficking problem.

2 Human Trafficking

Human Trafficking is a complex, global issue that affects millions of people worldwide each year, and as such requires a similarly global and comprehensive response in order to combat its continual spread [5]. Article 3 of the United Nations Protocol to Prevent, Suppress and Punish Trafficking in Persons provides the first agreed definition of human trafficking:

"Trafficking in persons" shall mean the recruitment, transportation, transfer, harboring 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. Exploitation shall include, at a minimum, the exploitation of the prostitution of others or other forms of sexual exploitation, forced labor or services, slavery or practices similar to slavery, servitude or the removal of organs' [6].

Human trafficking is a global criminal business that impacts on every country in the world. It is estimated to have a global worth of $32 billion and is recognized as a high profit, low risk crime. In 2012, The United Kingdom's National Crime Agency reported a 9% in the number of trafficking victims it had identified, with the same report indicating there had been a 130% growth of individuals being illegally trafficked to the UK for Cannabis Cultivation purposes [7]. Frequently in these cases the entire trafficking process is controlled by criminal organizations aiming to further exploit trafficked victims through inducting them into a variety of other illicit activities, such as forced labor and prostitution [7], [8]. EUROPOL and UNODC have also published statistics indicating that human trafficking is one of the most prominent and complex security issues across the globe today. [8], [9].

In order to improve the human trafficking defense architecture; data acquisition, application, and trans-border co-operation must be significantly improved, with a view to fostering an environment where knowledge is a key aspect of LEA's arsenal in combating perpetrators of trafficking [5]. To be truly effective, human trafficking identification must take a multidisciplinary approach, triangulating the requirements to maximize indicator identification, validation and perpetrator reprehension [10]. However, human trafficking indicators are often well concealed, with traffickers going to great lengths to obscure their activities, hiding the true size and nature of the trafficking problem [11], [12], [13]. The circumstances under which it takes place are also often complex and significantly varied from case to case [5]. Human trafficking, and particularly trafficking for the purposes of forced labor often involves victims that are considered vulnerable due to a variety of factors that include poverty, unemployment, inequality and political conflicts within their own states [7]. These indicators can be applied to strategically anticipate new destinations and transit routes taken by victims.

Human trafficking is traditionally considered to be under-reported, meaning that efforts to combat the problem cannot rely on observer reports alone [14]. The diverse range of potential criminal activity to which human trafficking is associated (i.e. money laundering, prostitution, forced labor and drug cultivation [15]) means that it provides an ideal use case upon which the application of KM, and specifically ontological knowledge representation and textual extraction through open-source scanning for combatting emergent organized criminal threats can be demonstrated.

In the UK, a number of recent LEA operations have uncovered instances where certain 'nail bars' have been used as fronts for trafficking activities [16]. The use of nail bars, and massage parlors as fronts for trafficking and the crimes to which it is associated is not solely isolated to the UK, with US law enforcement officials reporting similar issues [17], [18]. In the UK alone, ninety salons have been raided and prosecuted in the last five years for employing illegal migrants, concealing cannabis cultivation farms, prostitution rings and money laundering operations [19].

In response, open-source data mining has the ability to abstract data from official reports and news articles to provide a strategic outlook of the trafficking environment and combine it with operational level indicators derived from open social media to map potential hotspots of activity, and validate potential weak indicators of operational level activity. Further, the aggregation of operational data itself presents a strategic outlook of the current environment, enabling current indicators patterns ('primary' indicators) to be temporally extrapolated to indicate future or emergent threats [20].

The short narrative presented previously around the use of nail-bars as fronts for trafficking activity provides sufficient scope to demonstrate the potential capability of conceptual graphs, text analytics and the wider notion of KM in aiding the intelligence-led fight against human trafficking. As an emergent situation in the UK, the scenario described provides sufficient resource in terms of media reports to identify potential trends at a strategic level - such as patterns in migration, or hotspots of drug seizures at specific locations that may indicate trafficking. In addition, pages and posts indicating references of nail bars, and unusual activity and behaviors at the locations derived from the media information can also be applied to identify operational indicators of trafficking, and organized crime.

3 Knowledge Management

To ensure that the knowledge derived from open-source is utilized effectively we propose the use of KM, as a mechanism for LEA's to acquire, apply and store knowledge in a more efficient manner. Knowledge has increasingly become a key strategic resource for LEA's. The September 11th 2001 attacks in New York and Virginia simultaneously identified inadequacies in the current operations of public safety organizations, and the future requirement that knowledge be considered a key determinant of organizational success [21], [22]. Taking this rationale into consideration it is clear that knowledge has become a key strategic resource that has a tangible and quantifiable value in the intelligence-led fight against organized criminal and terrorist acts [23]. More generally, KM aims to gainfully exploit the intellectual capital of the individuals and organizations where it is being employed through the "systematic and explicit management" of all knowledge related activities including the practices, programs and policies that are set and followed by the organization [24]. The various disciplines that are holistically referred to as KM effectively exist across three domains; the organization (its structure, processes and controlling mechanisms), its culture (the informal, tacit practices of its individuals and occupational groups), and its technology (IT based systems) [25, pp.7-8]. Only through the effective management of all three of these domains can knowledge be leveraged effectively by decision makers.

Within the public sector, LEA's are increasingly acknowledging that due to the trans-national nature of organized crime, and in particular those of a trafficking related nature, there is a requirement for a coordinated, multi-agency approach to tackling said offences. UNODC have developed an intelligence sharing system in collaboration with global LEA's in order to facilitate intelligence sharing across borders and jurisdiction in order to aid in the fight against the growing threat of human trafficking. The resulting system, called the VRS-MSRC (Voluntary Reporting System on Migrant, Smuggling and Related Conduct), aims to provide an online outlet for LEA's to create and manage strategic knowledge that can be applied in the formulation of regional and national trafficking policies, informed by data from across global agencies through the analysis of routes used, transportation methods, financial transactions, victim and offender profiles, and other potentially relevant information [26].

Within the applied context of what they describe as 'Intelligence Management', Akhgar & Yates [23] define KM as a process of creating value added learning processes so that knowledge becomes a key strategic resource of LEAs with 'measurable and quantifiable value' in successfully combatting organized crime or an act of terrorism. In this view, in order to enable the understanding of dynamic, ever evolving real world problems a KM approach based upon a number of value added learning processes is required to enable a comprehensive understanding of the problem situation at hand. This approach is firmly grounded in epistemology, and places value on both human and machine learning processes, acknowledging the potential added value of human, tacit inquisition in strategic intelligence.

There is some debate among KM scholars however as to whether it is possible to abstract the tacit, un-codified aspects of human knowledge at all, with some (see [27]) arguing that in order to retain the contextual value of tacit knowledge, it cannot be codified or 'converted' into an explicit, expressible format, such as what is exemplified in Nonaka, Toyama & Konno's [28] SECI knowledge spiral. Boisot's model [29], named the i-Space suggests that in order to be converted, and disseminated, tacit knowledge must be codified and stripped of some of its contextual value in order to be recorded, and further disseminated explicitly. This poses a potential point of issue for ontology driven knowledge representation and other technological systems that make claim to being focused on, or influenced by KM.

Conceptual approaches to KM, such as the knowledge cycles proposed by Dalkir [25], Meyer & Zack [30] and Wiig [31], capture and model the lifecycle of knowledge assets within an organization from their acquisition right through to their application in aiding decision makers and solving problems. Within this, the technologically oriented aspect of KM most commonly targets the capture, creation and codification of knowledge [25, pp.77-97]. This 'explicit' knowledge must be formally expressible through the use of formal, systematic language so that it can be tangibly stored and shared through written mediums such as conceptual graphs and more commonly through items such as training manuals, codes of practice or mathematical formulae [28], [32]. Ontologies and the use of conceptual graphs provide one such approach to knowledge codification.

4 Ontology driven Knowledge Representation

Ontologically driven knowledge frameworks are static by nature, and thus can only provide solutions to known problems, based on the domain knowledge and relationships that have been modelled within them. This limits any potential impact of any ontologically driven solutions, as although the frameworks themselves are static, the environments upon which they are based are dynamic and ever evolving [23]. However, despite this and the earlier discussion around knowledge conversion, the role of ontologically underpinned knowledge cannot be ignored due to the requirement for technological systems to capture, model, organize and retrieve the constituent, abstracted parts of knowledge - i.e. act as a repository for codified organizational domain knowledge. Sowa [33] addresses this point of issue to some extent through emphasizing the dynamic and changing nature of information, and the complexities that these attributes cause for computer systems.

As a result, Technology and IT systems have to be constructed in a way that facilitates their evolution along with the environment. Increasingly, modern initiatives towards intelligence-led policing practice have also sought to integrate information systems within intelligence strategies. In the past these systems have comprised of what we may now consider to be simplistic document management systems and databases of existing policing records. Approaches such as that offered by IBM's 'COPLINK' [34], [35] have enabled LEA's to share and analyze vast quantities of existing records and information from across the intelligence community to aid national and border security. More recently, contemporary intelligence-led approaches have begun to integrate data analytics in intelligence-led activity for the detection, and predication of potential criminal activity based not only on historical crime records but also on emergent indicators in the environment. This is the mandate of the ePOOLICE project.

Conceptual graphs provide one such environment for ontology driven knowledge representation, providing a means to represent knowledge that is both understandable by software and people [36]. Ontology driven knowledge representation, and specifically conceptual graphs deal with data acquisition and KM though the use of an ontological vocabulary, and the subsequent relationships that are defined between the concepts in that vocabulary, providing a means to explore and model the semantics of domain knowledge [37].

4.1 Applying CGs using CoGui

Conceptual graphs (CGs) provide a valuable research tool for identifying and expressing the relationships between organized crime indicators, such as those defined by the International Labor Office Delphi survey [38], and the United Nations anti-human trafficking manual [10], and the behaviors and activity that can be identified through crawling other open sources and the text analytics that underpin it. One such tool for representing domain knowledge through the use of conceptual graphs (CGs) is CoGui. Aimed at enabling the visual expression of knowledge in the form of CG's, CoGui is heavily based upon Sowa’s [37] work on conceptual structures and knowledge representation. CoGui enables the composition of a set of ontology powered CG's capable of representing assertions [39]. In this section CoGui will be used to demonstrate the utility and capability of conceptual graphs in enhancing the value of the ePOOLICE project. In representing the ontological relationships between entities, CG's enable the definition of 'rules', as shown in figure 1. Rules enable the visualization of implied knowledge using an if-then (hypothesis-conclusion) scenario. The following section revisits the scenario presented in section 2 around the use of nail bars to demonstrate the applicability of conceptual graphs. The rule presented in figure 1 hypothesizes a generalized migration pattern, potentially facilitated by organized crime groups.

Fig 1. Victim Route Rule

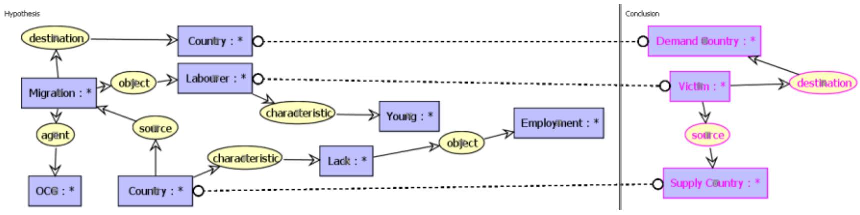
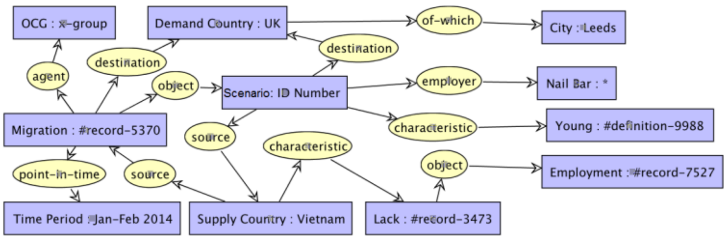


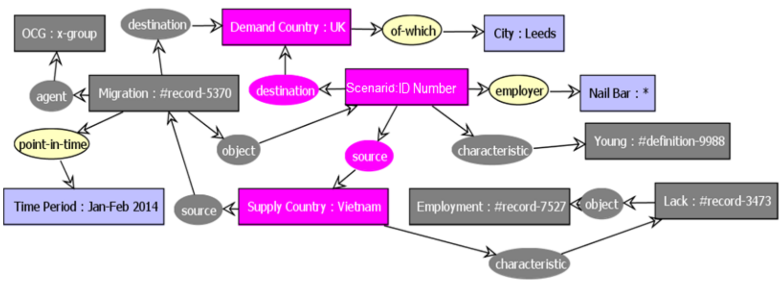
Figure 2 shows a fact CG that is derived from the textual extraction described in section 5. The CG identifies how the relationships between locations and nail-bars (the concepts extracted from open-source) are modeled to reveal their linkages with domain knowledge that exists behind the indicators. The fact CG represents a specific scenario, in this case migration movements between two specific locations, an event that is known to coincide with increases in trafficking.

Fig 2. Fact CG



The rule identified in figure 2 can then be applied to the fact in figure 1 to show the reasoning behind it. Figure 3 identifies through projection that the rule is indeed a potential instance of migration from Vietnam, with Leeds, UK the likely destination. The resulting fact graph is then used to update, and further enrich the wider knowledgebase. It is worth noting that in this particular example, although the activities highlighted, and relationships made between them potentially provide valuable information on events that may indicate trafficking activity, there is no evidence that suggests illegal activity is taking place in the case provided.

Fig 3. Output



In demonstrating the application of CG's using this simple example it is possible to see how the underlying concepts can applied as part of a larger knowledge repository to represent how indicators derived from open-source, using textual extraction and analysis techniques (such as those identified in section 5) can be abstracted to not only give an indication of operational trafficking, but also to provide a strategic perspective of migration patterns that may indicate the potential for a future, or emergent trafficking threat. In order to extract the indicators of such illicit activity from the open-sources, we introduce the use of textual analytics into the information retrieval process.

5 Textual Extraction and Categorization

Text analytics encompasses a wide variety or semantic and linguistic disciplines such as sentiment analysis, data mining, concept and contextual extraction, and content categorization [40] and is used to identify trends and previously unknown insights about their content [41]. Existing work in the area has focused on the extraction of named entities and concepts from unstructured closed-sources such as police case reports, enabling improved access to and co-ordination of such information [42], [43].One such aspect of text analytics; content categorization, is designed to create, maintain and apply semantic infrastructure to previously unstructured content [44]**.** In applying text analytics to open-source data it is possible to discern insight into groups and networks that was not previously possible due to the scale and level of access that provided by the web [45]. As human trafficking encompasses such a wide variety of potential issues, from forced labor, drug cultivation and distribution, the sexual exploitation of women and children, and numerous other forms of exploitation a taxonomy driven approach to categorizing disparate content enables decision makers to separate potentially relevant information from the noise of irrelevant open-source data.

Two approaches have been considered for categorizing content in this way; statistical and rule based. Statistical categorization relies on the acquisition of a robust corpus of training documents that accurately represent each content category, with documents matched to that particular category based upon their closeness to the statistical signature of the training corpus. The statistical signature is based upon measures such as the frequency at which particular words or phrases appear within documents. Although considered easy to develop, statistical models are entirely reliant on external content mined from the open-source environment being represented accurately by the training corpus. Despite having the benefit of being relatively simple to develop, the user has less direct control over the output, and thus the accuracy of a statistical approach tends to be lower [46, pp.201-202]. Due to the disparate nature of open-source data and the broad scope of the ePOOLICE project in categorizing data based on a wide variety of criteria and criminal activity, a rule-based approach consisting of both Boolean rules and linguistic terms is pursued. This approach enables the use of regular expressions, and linguistic and Boolean rules to define rules for each category in the defined taxonomy, thus enabling a higher degree of precision. Boolean rules although considered the most difficult and time intensive to develop, are also the easiest to refine and maintain, and potentially the most powerful and therefore accurate categorization method [46, pp.269-270]. In order to categorize content effectively, a representative taxonomy is first required based upon the variety and subject matter of the source material to be categorized as well as what the intended outcome of the categorization is [41].

5.1 Content Categorization

In order to establish a comprehensive understanding of the environment, it is first necessary to design a taxonomy that encompasses all domain relevant characteristics, in order to facilitate the categorization of content against these factors. For this purpose, a taxonomy based upon a number of existing publications including the International Labor Organization's 'Delphi' survey Indicators [38], UNODC's Anti-Human Trafficking manual for criminal justice practitioners [10] has been tailored to represent the specifics of identifying illicit nail bars in UK towns and cities. In this example (shown in Figure 4), documents related to mentions of young female workers will be categorized using Boolean syntax that uses logic stating that if a mention of an individual being 'young' or expressions that indicate youth appears in the same sentence, and within three words of syntax referencing a child or woman then the content is likely to be in reference to underage workers. Alone, this information does not necessarily provide any meaningful insights, however when content is similarly matched as being related to a location, or business such as a nail-bar then it provides a potential weak indicator of illicit activity.

Fig. 4. Young Workers Categorizer

```
(OR,
   (AND,
      (SENT,
         (DIST_3,
            (OR,"young","youthful","under-age","under age","under 16","under 18"
            ),
            (OR,
               (OR,"girl","girls"
               ),
               (OR,"child","children"
               ),
               (OR,"women","woman"
               ),
               (OR,"ladies","Ladies","Lady","lady"
               ),
               (OR,"she"
               )
            )
         )
      )
   )
)
```

5.2 Concept Extraction

A further application of text analytics is referred to as named-entity or concept extraction. In this instance, a list of location 'classifiers' are used in combination with Boolean rules in order extract explicit mentions of 'locations' in sentences that infer that something, or someone is transferred to or from it, the outcome of which could be applied to geographically map trafficking activity from reports and news articles to give a strategic view of the existing trafficking environment, and known routes.

Classifiers consist of a list of raw text concepts. In this instance these are made up of a list of countries and major towns and cities from across the globe extracted from MaxMind - a free database of major world cities [47].

Additional Concepts are then added to the taxonomy to identify the relationship between the 'Location' concepts and a particular use of language; in this case travelling to or from a destination. The syntax defined using Boolean rules states that if a term that indicates that someone/ or something is ‘arriving’ or ‘destined’ appears in a sentence with a concept from the ‘LOCATION’ rule then a positive match is returned to indicate a possible destination location, with the same logic applied to the OriginLocation rule. The ‘_c’ signifies the value to be returned, in this instance referring to the ‘LOCATION’ concept. The Boolean category rule qualifier ‘@’ is then used as a suffix in order to expand the search to contain all versions of the word forms it follows. For example, ‘arriving’ is expanded to include ‘arrive’, ‘arrived’ and ‘arrives’.

DestinationLocation Concept:

```
CONCEPT_RULE:(AND,(SENT,(DIST_4,(OR,"destined@","destinat
ion@","going@","arriving@","going@"),"_c{LOCATION}")))
```

OriginLocation Concept:

```
CONCEPT_RULE:(AND,(SENT,(DIST_4,(OR,"origin@","originates
@","source@","sourced@"),"_c{LOCATION}")))
```

Figure 5 shows these concept rules applied to a dummy ‘news report’ describing an artificial report of traffickers being apprehended in the UK city of Leeds. The blue-tagged data references the DestinationLocation concept rule, whereas the red-tagged data represents data matching the OriginLocation rule.

Fig. 5. Concept Rule Matches

Upon arriving in <DESTINATION_LOC><LOCATION>Leeds</LOCATION></DESTINATION_LOC> traffickers were apprehended by police. The nail salon workers were identified as being trafficking victims originating from <ORIGIN_LOC><LOCATION>Vietnam</LOCATION></ORIGIN_LOC>.

Best Matches

Concept	Matches
Top/LOCATION	2
Top/DESTINATION_LOC	1
Top/ORIGIN_LOC	1

Concept extraction, as demonstrated above can be applied in order extract known trafficking routes from textual data sources such as news articles and official reports and mapped geographically to show current, and historical trafficking hotspots, information which could be of potential benefit in validating weak indicators of emergent activity.

5.3 Weak Signals in the Environment

Further, similar syntax is applied to detect weak signals in the environment. The example in figure 6 demonstrates how tweets that are collated and analyzed from a micro-blog may provide indicators of labor exploitation, with nail-bars used as a front for such illicit activity. In this instance, after the content is filtered to remove information that may be used to identify individuals, the concepts and categorizer's defined in the taxonomy are used to analyze incoming data during its acquisition, with any rule-matches resulting the addition of the <Concepts> and <Categories> tags. The terms identified may aid in the detection of migration patterns, and as a result trafficking activity, due to appearance of new nail bars in one location. Further rules would then be defined to pick upon specific indicators of underage work, forced labor and sexual exploitation such as the mention of 'young looking girls' in figure 6.

Fig. 6. XML Document Markup


```xml
<?xml version="1.0" encoding="UTF-8"?>
<article>
<query>"Nail Salon" "nail bar"</query>
<authortimezone>London</authortimezone>
<doclang>en</doclang>
<body>The girls working at that new nail-bar in Leeds sure look young. I must ask them what their
secrets are for young looking skin!</body>
<LOCATION>Leeds</LOCATION>
<Categories>top\NailBar</Categories>
</article>

<?xml version="1.0" encoding="UTF-8"?>
<article>
<query>"Nail Salon" "nail bar"</query>
<authortimezone>London</authortimezone>
<doclang>en</doclang>
<body>Must have walked past 2 or 3 nail bars on the way to Elland Road this afternoon, they are
springing up all over the place!</body>
<LOCATION>Elland Road;Leeds</LOCATION>
<Categories>top\NailBar</Categories>
</article>
```


Although it is necessary to add further complexity, and comprehensiveness to the categorization and extraction rules presented to offer actual utility in the detection of organized crime signals in open-source data, it is clear that a developed version of such concepts can have real potential in the identification of tangible crime indicators in open-source data. Further, the underlying relationship between indicators and entities enables them to be generalized and extrapolated (i.e. codified and stripped of context) and applied to wider organized crime scenarios. In addition, the aggregation of operational 'primary' (i.e. current threat) indicators such as those identified previously can be used to map strategic trends and patterns in activity to identify emergent and future threats through the idea of temporal proximity [20].

6 Concluding Remarks

In this paper we have presented human trafficking as a potential use case for semantic knowledge representation in aiding to re-enforce both the strategic and operational capability of law enforcement agencies in combatting organized crime threats as part of a wider knowledge management strategy. KM has been identified and rationalized as an enabler of more effective intelligence practice, enhancing law enforcement agencies capacity to integrate new intelligence streams such as ePOOLICE into existing practice through improving communication and information sharing channels, knowledge acquisition and retrieval, and knowledge co-ordination and availability. Conceptual graphs are used to demonstrate one way in which state-of-the-art technologies can stimulate more effective KM practice and as a result improve LEAs response to organized crime threats. In addition, the utility of applying text analytics in extracting valuable information from disparate information sources has been demonstrated, and the requirement for further research as part of the ePOOLICE project identified in order to expand the basic concepts outlined here in order to have real, tangible impact in extracting organized crime indicators from 'primary' open-source data, and ultimately in informing domain decision makers.

Disclaimer

In this document, terms indicating origin or ethnicity are not being used to imply anything general or stereotypical of that origin or ethnic group. Such terms are only used as instances of factual reporting and are not be taken as a reference to any race or ethnic group as a whole. Nevertheless, for a system to monitor organized crime to operate effectively, the identification of certain elements such as gender, nationality and ethnicity, in addition to the explicit identification of crime gangs, victim groups and modus operandi are often important. For instance, Vietnamese victims are trafficked into Europe (supported by several reliable sources) and, therefore, the reference to the Vietnamese origin of criminal groups is crucial to investigate such cases as it forms a key characteristic of the phenomena described. Furthermore, when sensitive or personal data is being handled, it will be done so in accordance with laws protecting the privacy and human rights of individuals, including data protection laws.

Acknowledgement

This research has received funding from European Union Seventh Framework Programme FP7/2007 - 2013 under grant agreement nº FP7-SEC-2012-312651.

References

1. A. Elliott and J. Urry, *Mobile lives*, Routledge, 2010.
2. R. Krikorian. *Twitter Engineering Blog: New Tweets per Second Records, and How!* [Online]. available: https://blog.twitter.com/2013/new-tweets-per-second-record-and-how.
3. ePOOLICE. *ePOOLICE - About* [Online]. available: https://www.epoolice.eu/EPOOLICE/about.jsp.
4. P. Gottschalk, "Knowledge management technology for organized crime risk assessment," *Inf.Syst.Front.*, vol. 12, no. 3, pp. 267-275, 07/01 2010. http://dx.doi.org/10.1007/s10796-009-9178-8.
5. G. Rankin and N. Kinsella, "Human trafficking–The importance of knowledge information exchange," in *Intelligence Management,* Anonymous : Springer, 2011, pp. 159-180.
6. United Nations Office on Drugs and Crime (UNODC), "United Nations Convention Against Transnational Organised Crime and the Protocols Thereto," 2004.
7. Serious Organised Crime Agency (SOCA), "UKHTC: A Strategic Assessment on the Nature and Scale of Human Trafficking in 2012," 2013.

8. Europol. *Organised Crime Threat Assessment* [Online]. available: https://www.europol.europa.eu/sites/default/files/publications/octa2011.pdf.
9. United Nations Office on Drugs and Crime (UNODC), "Global Report on Trafficking in Persons," 2012.
10. United Nations Office on Drugs and Crime (UNODC), "Anti-Human Trafficking Manual for Criminal Justice Practitioners," 2009.
11. T.K. Logan, R. Walker and G. Hunt, "Understanding human trafficking in the united states," *Trauma.Violence Abuse*, vol. 10, no. 1, pp. 3-30, Jan 2009.
12. F. Laczko, "Enhancing Data Collection and Research on Trafficking in Persons," in *Measuring Human Trafficking*, Anonymous : Springer, 2007, pp. 37-44.
13. G. Tyldum and A. Brunovskis, "Describing the Unobserved: Methodological Challenges in Empirical Studies on Human Trafficking," in *Data and Research on Human Trafficking: A Global Survey* , F. Laczko and E. Godziak, Eds. Geneva, Switzerland: IOM International Organisation for Migration;, 2005.
14. United Nations Office on Drugs and Crime (UNODC), "Global Report on Trafficking in Persons," 2009.
15. EUROPOL. *EU Serious and Organised Crime Threat Assessment* [Online]. available: https://www.europol.europa.eu/sites/default/files/publications/socta2013.pdf.
16. Human Trafficking Foundation. *Modern Day Slavery in British Nail Bars* [Online]. available: http://www.humantraffickingfoundation.org/news/2013/modern-day-slavery-british-nail-bars.
17. P. Martin. *Nail Salons and Human Trafficking* [Online]. available: http://www.huffingtonpost.com/phillip-martin/nail-salons-and-human-tra_b_669076.html.
18. A. Siskin and L.S. Wyler, *Trafficking in persons: US policy and issues for congress*, Springer, 2012.
19. D. Arbuthnott, "Beauty and the beasts: Slaves in britain," *The Sunday Times Magazine*, 08/18/2010 2013.
20. Criminal Intelligence Service Canada (CISC). *Strategic Early Warning for Criminal Intelligence* [Online]. available: http://www.cisc.gc.ca/products_services/sentinel/document/early_warning_methodology_e.pdf.
21. T.D. Chaves, M.R. Pendleton and C.J. Bueerman, "Knowledge management in policing," *Community Oriented Policing Services (COPS) US Department Of Justice* 2005.
22. R.G. Hughes and K. Stoddart, "Hope and fear: Intelligence and the future of global security a decade after 9/11," *Intelligence and National Security*, vol. 27, no. 5, pp. 625-652, 10/01; 2013/11 2012. http://dx.doi.org.lcproxy.shu.ac.uk/10.1080/02684527.2012.708518.
23. B. Akhgar and S. Yates, *Intelligence Management: Knowledge Driven Frameworks for Combating Terrorism and Organized Crime*, Springer, 2011.
24. K.M. Wiig, "Knowledge management: An emerging discipline rooted in a long history," *Knowledge horizons: the present and the promise of knowledge management*, pp. 3-26 2000.
25. K. Dalkir, *Knowledge management in theory and practice*, Routledge, 2013.
26. United Nations Office on Drugs and Crime (UNODC). *Voluntary Reporting System on Migrant Smuggling and Related Conduct (VRS-MSRC)* [Online]. available: https://www.unodc.org/southeastasiaandpacific/en/vrs-msrc.html.
27. S.D. Cook and J.S. Brown, "Bridging epistemologies: The generative dance between organizational knowledge and organizational knowing," *Organization science*, vol. 10, no. 4, pp. 381-400 1999.

28. I. Nonaka, R. Toyama and N. Konno, "SECI, ba and leadership: A unified model of dynamic knowledge creation," *Long Range Plann.*, vol. 33, no. 1, pp. 5-34, 2/1 2000.
29. M. Boisot, "Knowledge assets: Securing competitive advantage in the knowledge economy," 1998.
30. M. Meyer and M. Zack, "The design and implementation of information products," *Sloan Management Review*, vol. 37, no. 3, pp. 43-59 1996.
31. K. Wiig, "Knowledge management foundations," *Thinking about Thinking.How People and Organizations Create, Represent, and Use Knowledge.Arlington (TX, USA)* 1993.
32. I. Nonaka and H. Takeuchi, *The knowledge-creating company: how Japanese companies create the dynamics of innovation*, New York: Oxford University Press, 1995.
33. J.F. Sowa, "The challenge of knowledge soup," *Research trends in science, technology and mathematics education*, pp. 55-90 2006.
34. H. Chen, D. Zeng, H. Atabakhsh, W. Wyzga and J. Schroeder, "COPLINK: Managing law enforcement data and knowledge," *Communications of the ACM*, vol. 46, no. 1, pp. 28-34 2003.
35. IBM. *IBM i2 COPLINK: Accelerating Law Enforcement* [Online]. available: http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?subtype=BR&infotype=PM&appname=SWGE_ZZ_ZZ_USEN&htmlfid=ZZB03016USEN&attachment=ZZB03016USEN.PDF.
36. S. Polovina, "An introduction to conceptual graphs," in *Conceptual Structures: Knowledge Architectures for Smart Applications,* Anonymous : Springer, 2007, pp. 1-14.
37. J.F. Sowa, *Conceptual structures: information processing in mind and machine*, Addison-Wesley Longman Publishing Co., Inc., 1984.
38. ILO (International Labour Office), "Operational Indicators of Traffcking in Human Beings," 2009.
39. CoGui. *CoGui in a Nutshell* [Online]. available: http://www.lirmm.fr/cogui/nutshell.php.
40. R. Feldman and J. Sanger, *The text mining handbook: advanced approaches in analyzing unstructured data*, Cambridge University Press, 2007.
41. T. Reamy, "Enterprise content categorization - how to sucessfully choose, develop and implement a semantic strategy," *Knowledge Architecture Professional Services (KAPS Group)* 2010.
42. M. Chau, J.J. Xu and H. Chen, "Extracting meaningful entities from police narrative reports," in *Proceedings of the 2002 annual national conference on Digital government research*, 2002, pp. 1-5.
43. H. Chen, W. Chung, J.J. Xu, G. Wang, Y. Qin and M. Chau, "Crime data mining: A general framework and some examples," *Computer*, vol. 37, no. 4, pp. 50-56 2004.
44. T. Reamy, "Enterprise content categorization - the business strategy for a semantic infrastructure," *Knowledge Architecture Professional Services (KAPS Group)* 2010.
45. X. Hu and H. Liu, "Text analytics in social media," in *Mining Text Data*, C.C. Aggarwal and C. Zhai, Eds: Springer, 2012, pp. 385-414.
46. SAS Institute Inc. *SAS Content Categorization Studio 12.1: User's Guide* [Online]. available: http://support.sas.com/documentation/onlinedoc/ccs/12.1/ccsug.pdf.
47. MaxMind. *Free World Cities Database* [Online]. available: http://www.maxmind.com/en/worldcities.