Using Knowledge Management to Assist in Identifying Human Sex Trafficking

Marisa Hultgren
San Diego State University
marisahultgren@gmail.com

John Persano
San Diego State University
persano@hsec.sdsu.edu

Murray E. Jennex
San Diego State University
mjennex@mail.sdsu.edu

Cezar Ornatowski
San Diego State University
ornat@mail.sdsu.edu

Abstract

The rise of technology has enabled traffickers to lure victims and sell them at a faster rate than ever before. However, the rise of the Internet has also created new avenues for prosecution and criminal investigations for law enforcement. This study uses a knowledge management approach of identifying an ontology of keywords that were then used to assess a sample of online adult ads to identify potential human trafficking victims. The value of this paper is in the production of a set of keywords that can be used by law enforcement to find and rescue human trafficking victims.

1. Introduction

The U.S. Government defines human trafficking as sex trafficking in which a commercial sex act is induced by force, fraud, or coercion, or in which the person induced to perform such act has not attained 18 years of age or the recruitment, harboring, transportation, provision, or obtaining of a person for labor or services, through the use of force, fraud, or coercion for the purpose of subjection to involuntary servitude, peonage, debt bondage, or slavery [31]. However, the Department of Homeland Security, DHS, has more recently shortened the definition of human trafficking to a modern-day form of slavery involving the illegal trade of people for exploitation or commercial gain [11]. The state of California’s Department of Justice (DOJ) has clarified the definition to human trafficking is a modern form of slavery that involves controlling a person through force, fraud, or coercion to exploit the victim for forced labor, sexual exploitation, or both [16]. While slightly different all three definitions are similar in context and for this paper we will use the California DOJ definition with the notation that there are two classes of human sex trafficking, those victims under 18 (minors) and those 18 or over.

Human trafficking is a serious problem with the state of California stating that 3 of every 1000 people are in some form of forced labor, including forced sex trade (CADOJ, 2015). This paper focuses on the sex trafficking aspect of human trafficking and proposes a method for detecting sex traffic based on analyzing online (Internet) ads. Why online ads? The rise of technology and the Internet has enabled traffickers to lure victims and sell them at a faster rate than ever before. However, the rise of the Internet has also created new avenues for prosecution and criminal investigations for law enforcement. Officials now have access to a vast amount of information about the sex industry. This study uses a knowledge management approach of identifying an ontology of keywords that can be used to assess online adult ads to identify potential human trafficking victims. The value of this paper is in the production of a set of keywords that can be used by law enforcement to find and rescue human trafficking victims.

This study is specific to California as it is a major hub for prostitution and sex trafficking. It encompasses many ports of entry and lies adjacent to the Mexican border region which is rife with human trafficking [14]. Additionally, California is the home of the researchers for this paper and is better understood by those researchers. The available evidence on California’s victim demography is varied. This is in part due to the clandestine nature of this industry which has led to divergent numbers across studies. According to the data that is available, victims in California tend to be women and children who are native to the United
States [16]. However, this accounts for fewer domestic individuals as compared to the national trend which is much higher [4].

The aim of this study was to comprehensively explore text based indicators of human trafficking in online classified ads using California and Backpage.com as a case study. The specific goals were to:

• Compile a list of terms and/or attributes which may indicate human trafficking in online classified ads
• Utilize this list to determine high risk postings and further indicators of trafficking
• Produce an analysis of California’s commercial sex industry utilizing information gathered from Backpage.com.

2. Background – Human Trafficking

Sex trafficking is a worldwide issue that affects both developing and developed nations. The U.S. is one of the top destinations for trafficking victims and in 2003 was second only to Germany [28]. An estimated 100,000 to 300,000 U.S. minors are at risk of sexual exploitation every year and between 2008 and 2010 U.S. task forces opened 2,515 cases of suspected human trafficking of which 82% involved sexual exploitation [12] [4].

California plays a significant role in the sex trafficking industry with the main markets in San Diego, Los Angeles, and San Francisco. An important geographic element in California is the southern border region with Mexico. Mexican border towns like Tijuana have large numbers of trafficked individuals that can be linked to migration [14]. Foreigners travelling north through Mexico and Mexican nationals migrating to large cities within Mexico or to the U.S. create a vulnerability among individuals, especially female populations, which in turn results in a high number of trafficked migrants [13]. Exacerbating this problem are Mexican cartels and criminal syndicates within the country that are increasingly turning to human trafficking as a money making scheme [10].

California lies right at the epicenter of the U.S. sex trafficking epidemic. It is one of the top five U.S. states that are particularly vulnerable to human trafficking because of the number of ports of entry, proximity to international borders, and migrant worker communities [9]. California plays a significant role in the sex trafficking industry with the main markets in large metropolitan cities, rural areas are also locations specific to sex trafficking [35]. Trafficking rings are very mobile and adept at setting up venues for prostitution in various out of the way locations [35].

The Polaris Project, a nonprofit committed to ending modern slavery, manages the National Human Trafficking Resource Center Hotline (NHTRC). They receive and record reports of potential victims of human trafficking. Although nearly every state had potential victims, California far outranked all other states including Washington D.C. in 2014 for tips about potential human trafficking. Of the California based tips in 2014, a little over 75 percent were cases specific to sex trafficking. Of the 685 cases of California based sex trafficking reported to NHTRC in 2014, 48 (roughly 5%) were from online ads. The biggest hotspots for tips were in San Diego, Los Angeles, the San Francisco/San Jose area, and Sacramento. Additionally, there are hotspots situated along Highway 99 through the center of California in Chico, Sacramento, Fresno, and Bakersfield [29] [30].

Technology is being used to fight and support human trafficking [16]. Police use methods such as individual analysis of photos to determine if the individual looked like a minor, suspicious telephone numbers, or “listings of persons suspected to be at risk for trafficking.” A study on closed human trafficking cases in four U.S. counties found that 27% of victims were first identified by law enforcement through the facilitation of the Internet [31]. Sites such as Craigslist and backpage.com were common venues for

California had 141 cases of suspected human trafficking in 2011, of which 76% were believed to be sex trafficking [16]. Between 2007 and 2012 there were 441 arrests made in California under Penal Code 236.1. – False Imprisonment and Human Trafficking, in which 113 resulted in convictions. The convictions occurred in the counties of Alameda, Los Angeles, Monterey, Riverside, Sacramento, San Diego, San Francisco, Santa Barbara, and Santa Clara. Of those convictions, 46 percent occurred in Alameda County (which includes Oakland) [16]. This does not constitute all instances of human trafficking convictions as the majority of cases are prosecuted under other laws such as pimping and pandering. These are often easier to prove and allow for longer sentences than convictions of sex trafficking [16].

In addition to large urban areas, sex trafficking also occurs in and around military bases, along major highways, and sometimes in rural areas. In San Diego some traffickers have been known to set up shop in military housing. They negotiate with a sailor for unlimited sex in order to work the base out of the individual's residence for a set period of time [10]. Although sex trafficking rings typically operate in large metropolitan cities, rural areas are also locations for sex trafficking [35]. Trafficking rings are very mobile and adept at setting up venues for prostitution in various out of the way locations [35].
intended to traffic her

Commonly, police investigations begin by law enforcement browsing through ads looking for girls with pictures that look very young. If the girl looks underage this may lead to the opening of an investigation [25]. Another approach is to compare an advertisement’s picture with the advertised age. If there is a perceived discrepancy with the girl looking younger than the posted then an investigation may also be opened. Although these techniques have been one way to locate potential victims, there are complications. First, it is extremely tedious and second, photos are not always accurate [25]. Many law enforcement agencies use these tactics with some success. However, as criminals become smarter, law enforcement tactics must become more sophisticated.

Trends in the use of victim’s photographs are changing. Recently recruited victims are less likely to have their picture used on advertisements. Pictures of other girls were used instead [5]. Sixtyfour percent of the time ads are likely to use alternate photographs when victims are trafficked online [5]. Many victims reported having to hide their face. Some said it was because they looked very young, while others said they did not know why, but perhaps it was so that people could not recognize them [5] [33]. Photos can also be doctored, even swapped out with a completely different girl than the one being advertised [25].

Criminals increasingly use online classified ads to sell victims. It is quicker than street prostitution and can add a level of anonymity. Social media sites have formed an endless, global supply of potential victims. Classified ads on sites such as Backpage.com and other Internet venues have made it easier than ever for pimps to connect with interested customers. The Internet sex trade alone brings in 1 billion dollars annually and it is ever growing. Although accurate numbers are difficult to generate, both the act of sex trafficking and the promotion of it online is unprecedented. This has resulted in the expedited exploitation of victims, but it also has created an opportunity for law enforcement to use the Internet as an investigative and prosecutorial tool.

The Internet is used as a platform for advertising sexual services. Traffickers either use conventional websites such as Backpage, Facebook, Instagram, Twitter, Craigslist etc. or they go underground using the “dark net”. Advertising online results in less risk to the trafficker because they can ensure before ever meeting the client that they are comfortable with the transaction and they can make the client come to them. In spite of this, there is a tradeoff in that the use of technology allows law enforcement to “track advertisements and set up stings.”

Trafficking occurs very rapidly, especially with the rise of the Internet. In June 2011 a woman from Texas escaped a man in San Jose who intended to traffic her for sex. Within less than 24 hours he had posted an ad online and secured a customer in San Francisco [19]. The Internet and the use of mobile devices have become a medium for traffickers finding clients and criminals looking for victims. The rise of mobile technology has allowed traffickers to be in constant communication with their victim. This results in a smooth flow from advertising of sexual services to communicating with the buyer and ultimately selling the victim.

Although technology has streamlined the process of trafficking, the use of mobile technology has also allowed for greater autonomy among victims. They often have access to mobile phones in order to communicate with their controller and potential clients. Additionally, many pimps allow their victims to post their own online advertisements. This is usually accompanied by strict limitations, but it has allowed for greater autonomy among trafficked victims than before. A study by the Thorn Foundation [5] interviewed 115 former victims of trafficking and found that 63% reported being advertised online with this being an increasing trend. The popularity of online ads has certainly increased and is still on the rise. This can also be seen in national revenue estimates from online prostitution ads which increased by 20 percent in only one year from 2010 to 2011. The use of Internet advertising has also led to an increase in victim autonomy. Findings in the 2015 Thorn [5] study suggest that traffickers are more likely to be the one to post ads on websites advertising services. However, 18% of respondents reported that they posted their own ads exclusively and 17% said that they and their controller both posted ads. Many victims who posted their own ads were required to post a certain number each day. One respondent reported copying and pasting the exact ad her trafficker wrote and posting it multiple times a day. One woman stated that she posted three times a day, morning, noon, and night. Although in the study women sometimes had access to the Internet it usually included strict parameters. They were only allowed to be on prostitution advertising sites and were only afforded enough time to post their ads, around 10-20 minutes. According to the head of a juvenile safe house in Arizona, over half of the girls that come to them were

advertisement (Farrall, 2014, p.57). This percentage can be increased through constant monitoring of known prostitution websites in addition to those mentioned.

Law enforcement use online classified ads to find low hanging fruit. Often they look through Adult Service ads to find pictures of people that look as though they are being trafficked [26].
trafficked on Backpage.com [17]. This number may have increased now with the closure of Craigslist’s erotic services in 2010.

Backpage.com is the leading U.S. website for online adult solicitation ads [21]. It accounts for 70% of all online commercial sex ads and also has the most user traffic [17] thus it was chosen as the best website for this study. Half of the respondents in the 2015 Thorn [5] study reported that they were advertised on backpage.com. This was by far the most common site for their postings second only to craigslist which has now shut down their Adult Service section in 2010.

Although prostitution is illegal in California both trafficked and non-trafficked individuals post ads for companionship on sites such as Backpage.com. Often the ads include a disclaimer for law enforcement. Backpage currently filters ads through the use of keywords then individuals look at every ad that is posted [17]. Considering the fact that such a substantial amount of girls are still trafficked on Backpage everyday it is evident many fall through the cracks. So although efforts are in place to automate the process of finding potential victims, there is still substantial room for improvement. Additionally, increased cooperation with law enforcement can be developed to include procedures for rapid reporting.

3. Characteristics Of Sex Trafficking Ads

Building a framework of sex trafficking indicators in online escort service ads provides a foundation for extracting keywords/ontology that can then be used by a constant web crawling system to identify and locate possible victims of trafficking. This section discusses the literature that explores traits common in cases of sex trafficking. These characteristics are used to build a framework of sex trafficking keywords/ontology for use in conducting this study.

Movement between Cities - Ads with evidence of frequent movement, or transience, may be a sign that the sex worker is a victim of trafficking. Pimps often move victims around from city to city in order to avoid law enforcement and to maintain control by keeping victims from building a social support system or becoming too familiar with a particular area [10]. This keeps victims disoriented and ignorant of where to seek help [16]. Ultimately this steady movement leaves women and girls consistently vulnerable to those controlling them. A study on trafficking in Silicon Valley showed that movement around the Bay Area was a frequent tactic used by traffickers [20]. There is also a marketing side to the frequent movement between cities. It serves to maintain a “fresh” product line for clients [16]. In this way, “new product” is constantly circulated to entice the consumer [21]. International women are often moved from the East to the West Coast, South to Northeast, and from urban to rural and vice versa. American women are trafficked across city and state borders as well as internationally [35]. Additionally, traffickers may post ads in other cities to gauge the market in that area based on the amount of hits they get [10]. Identification of area code networks on online Escort service ads can help inform these circuit patterns.

Ethnicity and National Origin - Ibanez [21] referenced the denotation of ethnicity as a possible sign of trafficking. There are a few things to qualify when including this as an indicator in our study because of the fact that an estimated 72% of sex trafficking victims in California are American citizens and being “American” is not an ethnicity and not necessarily a reflection of national origin.

Restricted Movement - Law enforcement officials believe that pimps consider it safer to conduct activities online utilizing incalls only [10]. Previous studies have used restricted movement seen through terms like “Incalls only” as a possible indicator of sex trafficking [21]. To control their victims, perpetrators often deny them their freedom of movement and keep them isolated.

Unconventional Sex - One ad that was found to be a victim of sex trafficking included the term “open minded” [33]. This may be a code word stating that unconventional and sadistic types of sex acts are allowed, therefore the likelihood of the poster being a victim of trafficking could be higher [37].

Minors Trafficked Online - Buyers who are interested in purchasing a young girl (often a minor) are solicited using certain keywords that can indicate whether a victim is being trafficked. Ads for underage victims sometimes use words such as: fresh, fresh meat, young, virgin, prime, coochie (shaved), non-pro, new, new in town, barely legal/18, college student/girl, lovely, daddy’s little girl, sweet, 1986 Firebird, new in the life, liked girls, youthful, and fantasy [5] [6] [27].

Phone Numbers and Area Codes - Phone numbers are often included in classified ads. Sometimes they are used by johns to look up reviews on a particular girl as alternate names or pseudonyms are used [21]. Often there are multiple phones used by traffickers and victims. Contract phones and disposable phones are both used in sex trafficking operations. Latonero [26] looked at the distribution of phone numbers using a simple Google search, the phone number of one suspected victim was found on seven other escort service sites. It was also found on the Backpage site of Los Angeles, San Francisco, and Las Vegas. Additionally, the same phone number was found on MyRedBook with a different name but
similar photographs. This widespread advertisement across the Internet and geographic regions indicates that this is more than just a lone prostitute out to make a few dollars. It could be an indicator of a sex trafficking ring. Information from area codes gives a strong indication as to the movement of victims and traffickers. The area code network offers both a source and destination. Maps can be generated to determine suspected routes being used to transport victims from one location to another.

4. Knowledge Management Ontology/ Keywords

The previous section discussed the literature base for establishing the different sources of potential keywords/ontology that can be used by a system for detecting potential human sex trafficking victims. This section establishes the literature and theoretical base for using a knowledge management system, KMS, as an implementation to assist in searching for human sex trafficking victims.

Davenport and Prusak (1998) view knowledge as an evolving mix of framed experience, values, contextual information and expert insight that provides a framework for evaluating and incorporating new experiences and information. They found that in organizations, knowledge often becomes embedded in documents or repositories and in organizational routines, processes, practices, and norms. They also say that for knowledge to have value it must include the human additions of context, culture, experience, and interpretation. Polanyi [34] and Nonaka and Takeuchi [32] describe two types of knowledge, tacit and explicit. Tacit knowledge is that which is understood within a knower’s mind and which cannot be directly expressed by data or knowledge representations and is commonly understood as unstructured knowledge. Explicit knowledge on the other hand is that knowledge which can be directly expressed by knowledge representations and is commonly known as structured knowledge.

Jennex [22] defines KM as the practice of selectively applying knowledge from previous experiences of decision-making to current and future decision making activities with the express purpose of improving the organization’s effectiveness. Also, Jennex [22] viewed a KM system, KMS, as that system created to facilitate the capture, storage, retrieval, transfer, and reuse of knowledge. The perception of KM and KMS is that they holistically combine organizational and technical solutions to achieve the goals of knowledge retention and reuse to ultimately improve organizational and individual decision making. This is a Churchman [8] view of KM that allows KMS to take whatever form necessary to accomplish these goals. Another key definition of KM includes Holsapple and Joshi [18] who consider KM as an entity’s systematic and deliberate efforts to expand, cultivate, and apply available knowledge in ways that add value to the entity, in the sense of positive results in accomplishing its objectives or fulfilling its purpose. Finally, [1] concluded that KM involves distinct but interdependent processes of knowledge creation, knowledge storage and retrieval, knowledge transfer, and knowledge application.

Ontologies codify knowledge by providing a simplified and explicit specification of a phenomenon that we desire to represent [15]. Ontologies are useful because they explicate components that define a phenomenon and, thus, can help in systematically understanding or modeling that phenomenon [18]. Keywords are specific examples from an ontology with a complete set of keywords defining an ontology.

Alavi and Leidner [1, p. 114] defined a KMS as “IT (Information Technology)-based systems developed to support and enhance the organizational processes of knowledge creation, storage/retrieval, transfer, and application.” Many researchers include ontologies in the KMS as a method for organizing and retrieving knowledge from the KMS [18] [2] [24] [36] [3], [23]. This research attempts to create an ontology by identifying keywords from the literature that describes the phenomenon of human sex trafficking. The ontology will then be used in a KMS that is first used in this research for attempting to identify victims of human sex trafficking in online advertising and then for aiding law enforcement to do the same.

5. Methodology

This is a systems development research methodology perspective on Action Research as described by Burstein and Gregor [7]. The basic elements of this method is to create an artifact that demonstrates the point of interest, for this paper creating an KM ontology/keyword framework that can be applied in a KMS to identify potential victims of human sex trafficking in online ads. The basic steps of this method ha the basic steps of using the literature/theory to establish a foundation for the research, constructing an artifact to perform the research, and then evaluating/reflecting on the use of the artifact. This section first describes how the literature review was performed as this literature formed the basis for creating the theoretical foundation for artifact construction. The rest of the section
describes how the artifact was created and then how data was gathered and used to evaluate the artifact.

To perform the literature review the following search terms were used in various combinations: “sex trafficking” “human trafficking” “United States” “America” “California” “Online” “Internet” “social media” “backpage” “craigslist” “classifieds” and “technology”. The literature review began by using Google Scholar to find articles that have been highly referenced and to establish a baseline of pertinent literature. Further searches were conducted using databases such as Ebscohost-Academic Search Premier, Scribd, and JSTOR. Additionally, the reference section of articles that particularly related to this study were used to find other similarly related articles. Often this led to the identification of further studies published by the same entity or individual that was relevant to the study.

Data for the study was gathered by scraping from the “Female Escort” section of Backpage.com February 11, 2015-February 16, 2015 for 15 different California cities/counties. The areas included in the study were Bakersfield, Chico, Fresno, Los Angeles, Merced, Oakland, Orange County, Sacramento, San Diego, San Francisco, San Jose, San Luis Obispo, Santa Barbara, Santa Cruz, and Stockton. There were some errors that occurred in the data scraping process. San Diego only produced about two and a half days worth of data due to an error, so the data sample is much smaller than that of other large cities. The data was converted into a .csv file then processed in a Microsoft Excel file. The data automatically generated attribute columns for each ad in the following categories:

1. Location of Ad Posting
2. Advertised Age
3. Text included in the ad
4. Phone number
5. Title of Ad
6. url
7. Date/Time

Further attribute information was extracted from each ad through an analysis of the text, title, and phone numbers. The url was used as an identifying factor for each unique advertisement. This study looked at 9 text based indicators of sex trafficking. The following list includes the percentage of total ads with the specified indicator present:

1. Duplicate Ad/Phone number used (54%)
2. Denotation of Ethnicity/National Origin (28.2%)
3. Unconventional Sex Advertised (13.4%)
4. Out of State Area Code (12.1%)
5. Disguised Phone Number (12.5%)
6. Restriction on Movement (11.8%)
7. Transient Language (9.0%)
8. Indications of Working in a Group (8.8%)
9. Minor Keyword Indicators (5.2%)

To ascertain which ads had potential victims of sex trafficking an artifact was constructed by extracting keywords from the data and categorized based on the previously listed indicators. To determine which ads had evidence of Advertised Unconventional Sex, Restricted Movement, Transient Language, Working in a Group, and Minor Sex Trafficking keywords were extracted from the text of each ad. To extract Ethnic/Nationality data, keywords were used within the filter feature of Microsoft Excel from both the text and title column. The Excel file was joined to an area code location table to determine the origin of the phone number. Area codes that were not represented on the table were individually searched online using Google. Area code information was used to determine possible movement of providers. 84% of the ads produced phone numbers, another 12.6% were manually extracted due to the phone number being disguised in some way. Often ads will put random characters between numbers, write out the word, or use letters rather than numbers. 3.5% had no phone number at all, these ads relied on either email addresses or the message system within Backpage.com.

6. Results

This study is an in depth look at the ontology of online classified ads and how it relates to sex trafficking. The goal of this study was to develop the framework for an automated system to search for online sex trafficking by using KM to identify an ontology (keywords) for that framework. Drawing on contextual research of common sex trafficking characteristics, a list of indicators was developed including Disguised Phone, Ethnicity/ Nationality, Transient Language, Group, Restricted Movement, Duplicate Phone/Ad, Minor Keyword Indicators, and an Out of State Area Code. This data was then used to analyze the geographic distribution of California’s commercial sex industry as it relates to the indicators of sex trafficking, the area code network of ads and the ethnic/national make-up of individuals advertised.

Keywords which indicate possible minor sex trafficking were compiled and analyzed against the advertised age. All ages past 30 were compressed for better visualization of the more prominent ages which are between 20 and 25. It is important to note that the advertised ages are not necessarily accurate, especially if a girl is being trafficked. It may be more likely that a victim of minor sex trafficking would be advertised somewhere around the age of 20-25. This is young enough to still attract the correct buyer, but old enough
not to raise law enforcement suspicion. The keywords used were compiled based on a number of studies [26], [30], [5], [19], [6], 5 and [27], 1. Counts were done on the scraped ads to determine how many times a keyword occurred and how many ads used multiple keywords. A number of terms were very prominent overall such as “sweet”, “young”, and “fantasy”. Some were used very rarely such as “Only 18”, “turned 18”, “very young”, and “little girl”. It is unlikely that traffickers would use such terminology for fear of attracting law enforcement. Terms such as “College student/girl”, “new in town”, and “fresh”, seem to be slightly skewed to the younger ages. This could be an indication of younger individuals. Many ads had multiple keywords. Ages less than 24 had the highest numbers of keywords. It was found that keyword indicators can be used in an analysis of online sex trafficking, but these terms must be consistently updated. It seems that the terms “College student/girl”, “new in town” and “fresh” are currently the best terms to use for identification of potential minor victims. This information should be further analyzed especially through the examination of ads with confirmed minor sex trafficking cases.

7. Discussion

All relevant information was analyzed to produce 9 indicators of sex trafficking. The overall distribution of each indicator throughout the entire study is shown in Figure 1. By far the most prominent indicator was “Duplicate Phone/Ad” (column 1 in figure 1). This is similar to the findings of a previous study in Los Angeles which showed that a small number of phone numbers accounted for a disproportionate amount of ads [26]. Surprisingly ethnicity/nationality (column 2 in figure 1) was fairly prominent, however, on reflection it is believed that this reflects the very diverse communities used in the study.

To determine the centrality of trafficking indicators, Figure 2 shows the prevalence of an indicator in ads that include 2 or more indicators and 4 or more indicators. The percentages in this chart are based only on the 4836 ads which had at least one indicator of sex trafficking. These numbers do not equal 100% because many ads had more than one indicator. The most common indicator, present in well over 50% of ads with at least 2 indicators, was that of Duplicate Phone/Ad. The next most prevalent indicator was the inclusion of an ethnicity or national origin. All other indicators hover at or below 15%. The exception to this is Minor Keyword Indicators which are lower relative to the other indicators. Figure 3 shows the number of ads with multiple key words, with 4836 ads having one indicator, 3126 had two indicators, 1354 had three indicators finally dropping to one ad with seven indicators.

It is evident that keywords can be used to develop indicators of human trafficking. Although Out of State Area Codes, Minor Keyword Indicators, and Unconventional Sex can be used to point toward possible cases of trafficking, Transient Language, Restricted Movement, and Disguised Phone numbers are more interconnected and can offer stronger indications of trafficking. The categories of Group, Ethnicity/Nationality and Duplicate Phone/Ad are most likely associated with massage parlors, but should be
included in a framework as each indicator is present in other forms of sex trafficking as well.

It is interesting that the number of ads with multiple indicators dropped quickly the more indicators there were. This was expected and illustrates that the process works. It was not expected that there were large numbers of human sex trafficking victims, and this study has found one very strong candidate which will be further discussed in the next section.

Backpage.com was used for this research as a case study but it only accounts for about 70% of all online escort ads. That still leaves 30% untouched; a number which may grow as law enforcement increase their scrutiny of the Backpage.com. This research has created a framework that can be implemented into various prostitution sites, law enforcement can constantly monitor for suspicious activity. This also expedites the process of investigation. Constant monitoring will alert law enforcement very quickly to potential victims and give them more time that would otherwise be spent manually checking ads. This is not to say that a web tool would completely replace physically looking at ads, but it can streamline the process by flagging those that are suspect.

8. A Deeper Look

There was one ad which included 7 out of 9 indicators of sex trafficking. By the time this study was completed the ad had been removed from Backpage.com. However, a simple Google search of the phone number and keyword “backpage” yielded another, older ad which had been posted by the same user. This ad was advertised in San Gabriel Valley and the advertised age was 19 rather than 23 which was the age given 2 months later when this study was conducted. A Google search of just the phone number yielded various other profiles of this individual including one on showescorts.com which provided details about the individual’s various identities. She mainly advertises in Los Angeles/San Gabriel Valley but also has ads in Orange County and Las Vegas. Additionally, she uses a Las Vegas phone number.

This escort ad had 7 of the 9 possible indicators of trafficking. Upon further scrutiny it was found that the individual exhibited other possible signs of trafficking that were not at first evident in the textual analysis. The individual used multiple ages, she has been advertised in two known sex trafficking hotspots, and she is African American which constituted 40% of the victims of federal sex trafficking cases between 2008 and 2010 [4]. This is an evidence of how data mining technology, this process could identify potential victims of trafficking and help prioritize investigations.

Although there is no way to be certain of whether this individual is truly a victim of sex trafficking, the goal of this study was not to definitively identify victims but rather to identify possible victims. Backpage.com only accounts for a certain number of prostitution ads. Even if the process of manually sifting through each ad produces meaningful intelligence, it is better to streamline this process in a way that can be molded to each new prostitution site. The first step to this process is the development of a framework.

Additionally, data from online Adult Service ads can be used to inform the study of the commercial sex industry and consequently, sex trafficking. Data garnered from this study showed indications of the geographic trafficking patterns across state borders and the demographic makeup of many individuals. This information can help identify previously unknown trends in sex trafficking.

9. Conclusions

As a central location for sex trafficking in the U.S., it is important that California lead the charge in innovative methods to combat this crime. California must continuously work to improve counter trafficking measures. This includes cyber safety education for children to decrease online recruitment of victims. There must also be further cooperation between government agencies, private businesses, and nonprofit organizations both domestically and internationally. Additionally, stronger legislation is needed to support the prosecution of criminals—both the purchaser and the trafficker. California has taken tremendous strides, many of which have been in these areas. But one of the strongest opportunities for further improvement in California is to leverage the vast amount of available online data as a tool to combat trafficking.

With criminals moving away from using real pictures in their online ads we conclude that other online tools are needed. This research provides evidence that an alert system based on a human sex trafficking ontology can be created. Although Latonero [26] concluded that such a system alone would not be effective, coupling a KM generated keyword based system with other KM practices may be effective. Using a KM approach to update the keyword ontology based on lessons learned from actual successes in rescuing trafficking victims will keep the basis for the system accurate. Periodically scraping ads and mining the results for additional keywords should also be done.
Further research must be done on how to automate the process of detecting inconsistencies in ages and aliases as well as third party language. Additionally, the use of PhotoDNA technologies should be bolstered across the conventional and dark web as this can greatly enhance the identification of possible victims. It is important that all of these tools are incorporated together. Otherwise, many victims will fall through the cracks.

Ultimately, the conclusion of this paper is that not using a technology based system to assist in identifying and rescuing trafficking victims is not acceptable. We have shown it is possible to use KM techniques to create ontological frameworks for analyzing for human sex trafficking. This research has demonstrated that an ontology can be created and ads analyzed for the keywords.

9.1 Limitations

This paper has one major limitation, it only looks at female victims. It is recognized that males can also be victims of human sex trafficking. Males were not included in the study as the study is a proof of concept study that is demonstrating if it is possible to use a KMS to identify victims. The subject of the study was limited to females to make it easier to design and demonstrate the KMS.

There is also a limitation in the paper itself due to length restrictions and a desire to keep the findings useful. More analysis was performed then is included in the paper due to the length restrictions. A list of the keywords used is not included because it is desired not to encourage a change in criminal behavior by publishing the list. The paper presentation will include more of the analysis and the list of keywords.

9.1 Areas of Future Research

There are two obvious areas of future research. The first is to research male victims to determine if the indicators and keywords are different from female victims. The second is to investigate technologies that can be used to create an automated KMS. Excel was used for convenience of demonstrating proof of concept but will not be sufficient for a large scale, automated tool.

10. References


