



Cómo citar el artículo

Beltrán Gómez, A. & Ordóñez Salinas, S. (2014). Sistema inteligente para la detección de diálogos con posibles contenidos pedofílicos. *Revista Virtual Universidad Católica del Norte*, 42, 164-181.

Recuperado de

<http://revistavirtual.ucn.edu.co/index.php/RevistaUCN/article/view/502/1041>

Sistema inteligente para la detección de diálogos con
posibles contenidos pedofílicos¹

Intelligent System for Detecting Dialogues with Possible
Pedophilic Contents

Système intelligent pour la détection de dialogues avec
possibles contenus pédophiles

¹: Grupos de investigación GITIS y GESDATOS, tesis de maestría Universidad Distrital Francisco José de Caldas, marzo 2010-enero 2012.



Adán Beltrán Gómez

Ingeniero de Sistemas, Fundación Universitaria San Martín
Licenciado de Matemáticas y Física, Universidad del Atlántico
Especialista de Ingeniería de Software, Universidad del Norte
Especialista de Matemáticas, Universidad del Norte
Magíster Ciencias de las Información y Telecomunicaciones Universidad Distrital Bogotá, Colombia
Investigador Universidad Manuela Beltrán, Bogotá, Colombia
Grupo de Investigación GITIS
adan.beltran@docentes.umb.edu.co

Sonia Ordóñez Salinas

Ingeniera de Sistemas Universidad Distrital Francisco José de Caldas
Estadística, Universidad Nacional de Colombia
Especialización Teleinformática, Universidad Francisco José de Caldas
Magíster en Ingeniería de Sistemas, Universidad Nacional de Colombia
Doctorada en Ingeniería de Sistemas y Computación, Universidad Nacional, Bogotá, Colombia
Docente investigadora Universidad Distrital Francisco José de Caldas, Bogotá, Colombia
Grupo de investigación GESDATOS
sordonez@udistrital.edu.co

Recibido:	8 de octubre de 2013
Evaluado:	28 de enero 2014
Aprobado:	2 de abril de 2014
Tipo de artículo:	Investigación e innovación

Resumen

El comportamiento pedofílico en internet presenta un crecimiento peligroso para la infancia. Los gobiernos en su interés por controlar este delito han usado muchas estrategias, dentro de las cuales se destacan dos: la primera es usar páginas ficticias que hacen el papel de señuelo para detectar usuarios pedófilos, y la segunda utilizar agentes de policía, que haciéndose pasar por niños logran atraer y engañar a los pedófilos para después arrestarlos, pero estas soluciones requieren de mucho personal calificado y generan problemas psicológicos en las personas que participan. Se presentan los resultados de investigación relacionados con la construcción de un clasificador automático utilizando técnicas de aprendizaje supervisado y de procesamiento de lenguaje natural, adicionalmente la implementación en un agente software para monitorear en forma automática las conversaciones con posible contenido pedófilo en salas de chat, como herramienta para las entidades encargadas de la protección de la infancia.

Palabras clave

Clasificación automática de textos, Clasificación automática de Conversaciones, Sistemas Inteligentes, Pedofilia en internet.

Abstract

Increasing Internet pedophile behavior is dangerous for childhood. Governments are interested in controlling this crime therefore they have used many strategies, among which we highlight the following: firstly using dummy pages serving as lure for detecting pedophiles users and secondly using police officers that impersonating children attract pedophiles and then arrest them, but these solutions require a lot of skilled personnel and generate psychological problems in involved people. We

present the research results related to the construction of an automatic classifier using supervised learning techniques and natural language processing, we also present the implementation in a software agent for automatically monitoring talks with possible pedophile content in chat rooms as a tool for agencies responsible for child protection

Keywords

Automatic text classification, Automatic conversation classification, Intelligent Systems, Internet pedophilia.

Résumé

Le comportement pédophile sur Internet présente une croissance qui est dangereux pour les enfants. Les gouvernements sont intéressés dans contrôler ce délite, par conséquent ils ont utilisée beaucoup de stratégies comme l'utilisation de pages web fictives comme leurre pour détecter les utilisateurs pédophiles et l'utilisation d'agents de police que, en supplantant aux enfants, attirent et trompent aux pédophiles pour les arrêter, mais ce type de solutions requièrent beaucoup de personnel qualifié et produisent des problèmes psychologiques chez les personnes qui participent. On présente les résultats de recherche liés à la construction d'un classificateur automatique en utilisant des techniques d'apprentissage supervisé et de traitement du langage naturel, et aussi l'implémentation d'un agent logiciel pour le monitoring automatique des conversations avec possible contenu pédophile sur les salles de chat comme un outil pour les organismes responsables de la protection des enfants.

Mots-clés

Classification automatique de textes, classification automatique de conversations, systèmes intelligentes, pédophilie sur Internet.

Introducción

La tecnología permite en la actualidad un gran flujo de información, facilita la comunicación, en especial internet ha permitido el acceso a todos los campos del conocimiento y popularizado la información y el contacto con usuarios de cualquier parte del mundo. En este contexto existen aplicaciones de comunicación conocidas como "salas de chat" que permiten la interacción y comunicación entre más de un usuario en tiempo real y sincrónico a través de mensajes escritos.

Las ventajas de este tipo de comunicación se hacen evidentes al poder mantener una conversación virtual (es decir sin una presencia física personal) entre personas ubicadas en lugares distintos, ya sea a nivel local, regional, internacional o global. Con ello se hace el llamado ahorro en los proceso comunicativos en recurso, tiempo y eliminando las barreras de las distancias; aumentado la eficiencia y la eficacia en el proceso comunicativo.

Una de las características de la comunicación a través del chat es la del anonimato del usuario. Esto en muchas ocasiones le permite desinhibirse verbalmente y poder dar un rumbo a la conversación en el sentido que quiera orientarla. En este amplio y desregulado universo de la palabra son muchos los usuarios mayores de edad los que encuentran el ambiente propicio para la consecución de propósitos, como el de contactar menores de edad y sostener diálogos de contenido sexual con el propósito de llegar a mantener relaciones sexuales con estos, es decir, estamos ante un abuso sexual como lo define Beate Besten².

En concordancia con lo anterior, existe la posibilidad que menores de edad entren a ciertas salas de chat y puedan verse expuestos a ser víctimas de personas que tengan algún tipo de trastorno psicológico como la pedofilia, que es diagnosticada clínicamente, según el Manual de Enfermedades Mentales, como un trastorno psicosexual que define a quien lo padece como un individuo que fantasea con, es atraído por, o experimenta compulsión sexual hacia los menores impúberes (menores de 13 años) y cuyas actividades se realizan por un periodo superior a los seis meses; y cuya diferencia en edad entre el pedófilo y su víctima es superior a cinco años de diferencia (Association & DSM-IV., 2000).

Se propone la implementación de un agente de software para identificar las conversaciones con contenido pedófilo que puedan presentarse en un chat, fundamentado en técnicas de procesamiento de lenguaje natural y algoritmos de aprendizaje automático supervisado.

En el desarrollo de la investigación se utilizan dos metodologías: una para el diseño del clasificador, basada en el desarrollo de sistemas inteligentes. Para el desarrollo del agente software se utiliza TROPOS una metodología orientada agentes.

Uno de los principales desafíos del presente estudio fue la construcción de un *corpus de conversaciones* de contenido pedófilo, que permitió entrenar el algoritmo utilizado para la identificación de diálogos con características erótico-sexual entre un adulto y un menor de edad y que lleven a concluir que se está frente a una conversación pedófila. Toda vez, que se presentan inconvenientes debido a los problemas que conlleva la violación de la privacidad.

La consecución de este tipo de colecciones se ve limitada debido por un lado a la posible violación de la privacidad y a que esta clase de investigación se desarrolla generalmente en forma privada por las instituciones de seguridad.

Si bien el lenguaje natural ofrece complejidad en el procesamiento automático, el lenguaje usado en las salas de chat agrega elementos que lo dificultan aún más. Toda vez que los morfemas (segmento lingüístico mínimo dotado de significado gramatical y léxico) pueden ser escritos total o parcialmente, con errores ortográficos, o creación caprichosa de los mismos por parte de los usuarios.

Planteamiento del problema

Las actividades de pedófilos se han incrementado con el aumento de aplicaciones para hacer contactos en internet, por la facilidad de relacionarse tanto con menores de edad como con otros pedófilos; varios proyectos e iniciativas para investigar este fenómeno se han desarrollado a nivel mundial (Breedon & Mulholland, 2006) (Patterson, s.f.).

². Abuso sexual es todo acto no fortuito consciente o inconsciente, violento, aunque no obligatoriamente físico y siempre psíquico que sirve exclusivamente para satisfacer las necesidades del adulto y que se practica ejerciendo un poder, psíquico o físico que ataca a la esfera sexual de los niños, y que se produce con mayor frecuencia en familias e instituciones, antes que en extraños; por lo general no es un hecho aislado, causa lesiones físicas y sobre todo psíquicas que con frecuencia influyen negativamente a lo largo de toda la vida, y que afecta y amenaza en un niño el bienestar y sobre todo los derechos, y que el niño identifica como abuso (Besten Beate. Abusos sexuales en los Niños. Herder. 2001. Segunda Edición. Barcelona. p.19).

Para resolver esta problemática se han utilizado diferentes estrategias: por un lado se han creado páginas web ficticias que hacen el papel de señuelo para usuarios pedófilos, como herramienta estratégica para identificarlos; de igual manera, agentes de policía y voluntarios, haciéndose pasar por niños en las salas de chat, tratan de atraer y engañar a los pedófilos para después arrestarlos (Durkin, 1997; Penna, Clark & Mohay, 2005).

La automatización del proceso de detección de actividades pedófilas en internet se ha hecho necesaria, ya que las estrategias utilizadas hasta el momento, aunque han dado como resultado varios arrestos conocidos internacionalmente (Apuzzo; Lunau; MacKean, 2004; Stone & Miller, 1999), presentan dos tipos de problemas: el primero es el excesivo número de agentes necesarios para llevar a cabo una buena actividad; y el segundo, que debido al carácter de engaño, se corre el riesgo de hacer que usuarios de internet se vean seducidos por los agentes de policía a actuar como pedófilos sin que realmente lo sean (Fulda, 2002), se habla de abuso de menores pero no de pedofilia.

En el caso colombiano las cifras que revela el Instituto Colombiano de Bienestar Familiar muestran que aproximadamente 35.000 niños son explotados sexualmente en el país (Hernández, 2008). Por ello las autoridades competentes realizan una labor de seguimiento basada casi que exclusivamente en las denuncias que presentan los ciudadanos pero se corrobora que hay una ausencia de acción preventiva por parte del Estado, que no ejerce control y vigilancia en este tipo de delitos pues carece de herramientas suficientes para su pronta identificación.

Es aquí donde la investigación e implementación de tecnologías como el procesamiento del lenguaje natural y agentes o sistemas inteligentes, dirigidas a combatir delitos como la pedofilia a través de Internet son necesarias ya que este tipo de delincuentes se apropia cada vez más del saber tecnológico y cibernético para lograr sus objetivos.

Justificación

En toda la Red se puede encontrar lo bueno, malo, útil, efímero, etc., de la información y gracias al anonimato del que gozan los usuarios en internet se facilita también la posibilidad de cometer los llamados delitos digitales o ciber-crímenes; ya que desde allí se puede encontrar desde un manual de terrorismo que enseñe paso a paso cómo fabricar una bomba hasta imágenes y vídeos relacionados con conductas sexuales.

En este sentido cabe resaltar que una de las conductas ilícitas más frecuentes es la pornografía infantil, y que según datos de UNICEF, esta va en aumento. Se conoce de pedófilos que usan las salas de chat en internet para buscar a sus víctimas. Se hacen estimaciones de que existen cerca de 26.000 pedófilos. Incluso hay algunos que son conocidos como productores de películas en las que se usan a niños y niñas. Estas mismas cifras señalan que más de un millón de niños son fotografiados o filmados por año, para satisfacer dicha demanda que en términos económicos genera entre dos a tres millones de dólares (Unicef, 2010).

La Policía Nacional de Colombia señala que la tasa de los delitos contra la libertad, integridad y formación sexuales en Colombia se ha incrementado en los últimos años (Espino-Duque, 2010). La información que da la Unidad de Delitos Informáticos de la misma entidad establece que el uso de internet permite a los niños contactarse con pedófilos fácilmente.

Pese a que Colombia todavía tiene un gran trayecto que recorrer en cuanto a proteger y garantizar los derechos de los niños y niñas, ha habido grandes avances en legislación y la Constitución Nacional consagró tales derechos como un derecho de carácter Superior. En materia de lucha contra la pornografía infantil se cuenta con la ley 679/01 que expide un estatuto para prevenir y contrarrestar la prostitución, la pornografía, el turismo sexual con menores. El Decreto 1524 de 2002 establece las medidas técnicas y administrativas destinadas a prevenir el acceso de menores de edad a cualquier modalidad de información pornográfica contenida en

Internet o en las distintas clases de redes informáticas a las cuales se tenga acceso mediante redes globales de información.

Un paso importante para la prevención de actividades pedófilas en internet es la implementación de un sistema para la detención de conversaciones con posible contenido pedófilo en salas de chat, realizado por sistemas que usan técnicas de procesamiento de lenguaje natural, cuyas aplicaciones entre muchas otras son detectar el tema de una conversación y clasificar automáticamente los textos, basándose en ejemplos para generar patrones (Romaniuk, 2000).

Desarrollo de los antecedentes

El análisis automático de conversaciones en los *chats* ha gozado de gran interés por parte de los organismos de seguridad y de los agentes protectores de los derechos de los niños para prevenir delitos relacionados con la pedofilia. Sin embargo existen algunas limitantes en dicha prevención, toda vez que es difícil el acceso a las conversaciones del *chat*, pues hay un derecho protegido por la ley que es el del respeto a la privacidad de las personas. Además estas conversaciones poseen características gramaticales únicas que no se rigen por las normas de la escritura o la oralidad propias de un lenguaje natural, sino que se mezclan palabras o letras con símbolos matemáticos, gestuales y aparecen siglas nuevas, apelativos, entre otras, por ejemplo se mezclan palabras de diferentes idiomas y generan unas propias, complejizando el análisis de estos textos (Zitzen & Stein, 2004).

La sociedad ha venido cambiando generación tras generación, cuestionando y rechazando ideas y comportamientos que antes eran aceptados sin ningún tipo de reflexión. La diferencia entre mayor y menor de edad resultó de la capacidad del sujeto por razones de madurez psicológica, sexual y de responsabilidad en general y definió los límites a las relaciones entre personas y específicamente diferenciar estos dos niveles del desarrollo humano.

Desde la segunda mitad del siglo XX han ido generándose iniciativas de protección de los niños de los abusos de que eran víctimas. En 1946 se creó la Unicef con el propósito de proteger a los niños en todo el mundo y promover acciones que lleven a tal fin (Burger et al., 2001). De esa época a hoy se ha visto un aumento en las iniciativas gubernamentales para brindarles protección a los niños, identificando y encarando las amenazas a las que se ven expuestos.

Las actividades pedófilas, aunque son definidas dentro de un marco médico y de salud mental, desde mediados de los años 80 también tienen una connotación legal que las define como delitos cuando se sobrepasan los límites para el contacto sexual entre un menor de edad y una persona que pueda tener alguna ventaja sobre él, ya sea por desigualdades en cuanto a la edad, la madurez, o el poder ejercido (Fernández, Fernández, Victoria, Garrido & María, 2008). En este sentido es que las autoridades policiales desarrollan actividades de seguimiento, identificación y detención de explotadores y agresores sexuales de niños.

El avance de las tecnologías de la información y las comunicaciones, ha permitido que las personas tengan una mayor facilidad de contacto, ventaja que ha sido aprovechada por delincuentes para llevar a cabo sus actividades. Con el desarrollo de Internet se ha visibilizado el problema del abuso sexual a niños. Los pedófilos han sacado provecho del anonimato que les brinda Internet en las páginas Web y en las salas de chat y la utilizan como herramienta para establecer relaciones con las posibles víctimas y con otros victimarios para llevar a cabo intercambio de pornografía infantil y para acercarse y acechar a sus potenciales víctimas.

Dentro de las actividades de los pedófilos se encuentra el *"grooming"*, término con el que se conoce el proceso mediante el cual los pedófilos preparan sus víctimas para asegurar su cooperación y reducir el riesgo de ser descubiertos, usando las salas de chat, correo electrónico, mensajería instantánea y otros tipos de software para crear relaciones de confianza y poder luego, acosar sexualmente a los niños, mediante intercambio de

materiales eróticos y pornográficos para desensibilizarlos poco a poco, llegar a contactarlos y a través de regalos manipularlos (Stratica, Kosseim & Desai, 2005).

Ante el creciente aumento de estos delitos, las organizaciones sociales dedicadas a la protección del menor y la policía han desarrollado herramientas para contrarrestar las conductas pedófilas en internet, mediante la creación de páginas web en las que intentan atraer a los pedófilos para así poder acercarse a ellos e identificarlos. De igual manera, voluntarios y agentes entrenados, haciéndose pasar por menores de 13 años, ingresan a las salas de chat donde logran hacer contacto con posibles abusadores de niños estableciendo una relación lo más confiable posible y sin dar a conocer su verdadera identidad concretan un encuentro físico con el potencial abusador, momento en el cual las autoridades proceden a capturarlo y procesarlo judicialmente (Penna et al., 2005).

Aunque estas actividades de defensa y prevención hayan dado resultados positivos, cabe resaltar el bajo porcentaje que estos arrestos representan con respecto al total de la actividad pedofílica que se lleva a cabo en internet, ya que mejores resultados demandarían mayores recursos físicos.

A continuación se mostraran los avances tecnológicos y teóricos que permitirán detectar conversaciones en salas de chat donde posiblemente exista un dialogo de contenido pedófilo.

Monitoreo y análisis de chat Rooms

Existen programas de monitoreo de transferencias de datos, denominados "sniffers" que permiten mostrar los paquetes de datos que son transferidos de un equipo a otro en una red. Estos paquetes pueden ser analizados para identificar el emisor y el receptor, como también el contenido de los datos transferidos. Existe también el servicio "whois", que permite encontrar información para rastrear una dirección IP (Internet Protocol) a través de los proveedores del servicio de Internet - ISP (Internet Service Provider).

Dentro de los sistemas propuestos, a nivel investigativo, para el análisis de contenidos en *chat rooms* se encuentra por ejemplo, el programa *Butterfly* que es un agente de software desarrollado para muestrear los grupos de conversación disponibles en un *chat room* (Neil W. Van Dyke). Está basado en el protocolo IRC (Internet Relay Chat), este agente se introduce en los *chat* por un periodo de 30 segundos en los cuales puede identificar la temática del canal con base a las palabras que los usuarios están empleando, a través de un modelo de representación vectorial de las palabras.

Otra herramienta propuesta a nivel investigativo, aunque actualmente ya no se encuentre disponible, para la detección y prevención de delitos en internet, es la llamada *Chat Track* (Bengel, Gauch, Mittur & Vijayaraghavan, 2004), que actúa detectando discusiones en los chat y clasificándolas en categorías ya definidas. Para ello, inicialmente está entrenado con 1500 categorías que son obtenidas de *Open Directory Project* (ODP)³, utilizando técnicas de clasificación automática basada en representación vectorial. Para hacer la clasificación el robot toma una representación vectorial de la conversación y la compara con las categorías ya entrenadas a través del criterio de similitud coseno.

Debido a la simultaneidad de las conversaciones que se producen en las salas de chat, y la complejidad de hacer un seguimiento a todos, con este fin se ha propuesto *PieSpy* (Copestake & Jones, 1990), un robot IRC, que tiene como propósito determinar las relaciones dentro de un conjunto de canales IRC de las salas de chat y mostrarlas visualmente. La heurística utilizada por esta herramienta permite determinar las relaciones entre dos usuarios a través de los tiempos o rapidez en el intercambio de mensajes entre los "Nickname" (apodo que utiliza el destinatario en la conversación)(Mutton, 2004).

³. Disponible en <http://www.dmoz.org/>

Algunos métodos adicionales para clasificar los perfiles de los usuarios de una sala de chat son ofrecidos por las Investigaciones de (Çamtepe, Krishnamoorthy & Yener, 2004). Ellos localizan dentro de la sala de chat los patrones de la comunicación usando el algoritmo computacional de semántica latente basado en el Valor Singular de Descomposición (SVD - por sus siglas en inglés). De esta manera se llega a determinar grupos de usuarios hasta con un 70 a 80% de precisión.

Por medio de métodos de análisis estadísticos y un modelo de sala de chat, (Çamtepe, Goldberg, ismail & Krishnamoorthy, 2005) han llevado a cabo una investigación experimental cuyo resultado fue un algoritmo que determina grupos de conversación en las salas de chat, sin usar un análisis lingüístico.

Otras investigaciones relacionadas han sido llevadas a cabo por (Baumes, Goldberg, Magdon-Ismail, & Wallace, 2004), quienes muestran que con el tiempo, es posible identificar grupos ocultos en la red utilizando su modelo de comunicación. Sus investigaciones consisten en demostrar que esto es posible sin recurrir al análisis del contenido de las conversaciones.

Por su parte (Chen, Goldberg, & Magdon-Ismail, 2004) usan un modelo de simulación de comunicaciones para demostrar su algoritmo basado en características estadísticas de los mensajes, identifican de manera efectiva los grupos de noticias y en una sala de chat a los usuarios que utilizan múltiples identificaciones para enviar mensajes.

Clasificación automática de conversaciones

Dentro de las técnicas más usadas para clasificar conversaciones están los modelos de SVM (Maquinas de Vectores de Soporte). En un estudio piloto realizado por (Pendar, 2007), se aplicó esta técnica para identificar pedófilos en línea y reportó modelos de SVM y K-NN. En este caso como criterio de similitud se usó distancia euclidiana y como peso de los atributos se utilizó frecuencia TF-IDF (*Term Frequency-Inverse Document Frequency*) alcanzando un valor de F (F -measure) de 0.943 en datos de prueba, distinguiendo pederastas y voluntarios posando como víctimas en las salas de chat, demostrando que es posible, dada una conversación, diferenciar los textos correspondientes al depredador de la víctima.

Por otro lado, (Kontostathis, Edwards, & Leatherman, 2009), desarrollaron un proyecto multidisciplinario para analizar, indexar y categorizar las estrategias de comunicación utilizadas por los pedófilos en internet para manipular a los menores de edad. Ese proyecto desarrolló la herramienta *ChatCoder* para automatizar mediante técnicas de minería de textos el etiquetado de conversaciones con contenido pedófilo.

Un método para la detección del tema de conversación en un chat es el propuesto por (Paige H Adams & Martell, 2008), basado en el modelo de espacio vectorial. Pondera los términos mediante TF-IDF, y considera la frecuencia del término en el post de la conversación y en la colección completa de conversaciones. Este modelo aplica el criterio de matriz de conectividad propuesta por (Wang, Joshi, Cohen, & Rosé, 2008) para establecer el enlace entre pares de mensajes usando el criterio de similitud de coseno.

Para finalizar, (Martin, Appelt, Grosz, & Pereira, 1986) presenta un *corpus* de conversaciones etiquetado automáticamente de *chats*, que utilizan el protocolo IRC (*Internet Relay Chat*), donde participan varias personas. Hace uso del algoritmo de Máxima-Entropía para desarrollar el clasificador y determinar si dadas dos sentencias, de un grupo de sentencias de un chat donde hablan varias personas, estas pertenecen al mismo hilo de conversación tomando como criterio de comparación los resultados de un corpus etiquetado manualmente y evidencia que es posible desambiguar los diferentes hilos de conversación teniendo en cuenta características del discurso y el contenido de la conversación.

Sistemas inteligentes y agentes

Aunque no existe una definición unificada de sistemas inteligentes, en el presente contexto se toma la definición de (Carranza & Rodríguez, 1997):

- Sistemas computacional o físico en el cual se utilizan técnicas de inteligencia artificial o computacional.
- Sistemas que resuelven problemas complejos mediante la utilización de una representación de los conocimientos teóricos, destrezas, habilidades y pericias humanas.
- Sistemas que poseen la capacidad de raciocinio, comprensión y aprendizaje para realizar tareas que exigen algún grado de inteligencia cuando las ejecuta un ser humano.
- Sistemas que modelan y analizan fenómenos o sistemas reales complejos y pueden operar con datos, información o conocimientos incompletos, inciertos y/o difusos.

En cuanto al término Agente, en el presente trabajo se entiende en el contexto de agente no autónomo (Russell, Norvig, Canny, Malik, & Edwards, 1995). El agente puede contar con un conocimiento inicial pero a medida que el agente adquiere experiencia el conocimiento puede modificarse y aumentar. Se dice entonces que un agente no es autónomo cuando depende del conocimiento dado inicialmente más que a las percepciones de este, a diferencia de un agente racional que debe ser autónomo, debe saber aprender a determinar cómo tiene que compensar el conocimiento incompleto inicial.

Propósito y la fundamentación

A partir del planteamiento del problema y la revisión de los antecedentes, se planteó el siguiente cuestionamiento para su desarrollo: se puede implementar una herramienta software que permita a través de agentes y clasificación automática detectar las posibles conversaciones con contenido pedófilo? Por lo tanto, se plantea que los ejemplos de conversaciones que contienen características pedófilas junto con aquellas que no lo tengan, permiten desarrollar un clasificador que será utilizado por un agente software, para detectar posibles nuevas conversaciones con este tipo de contenido, a través de técnicas de clasificación supervisada.

Metodología

Para la construcción del modelo de clasificación se evaluaron diferentes algoritmos de aprendizaje automático utilizando el corpus de conversaciones, construido para el desarrollo de este proyecto. Para la construcción de dicho corpus se aplicaron las técnicas necesarias de procesamiento de lenguaje natural. Las técnicas de clasificación se aplicaron teniendo en cuenta algunas variaciones como por ejemplo el incluir o no técnicas de *stemming*, frecuencias absolutas, TF-IDF, entre otras. Con el fin de evaluar los diferentes modelos de clasificación se utilizaron medidas como el índice de recuperación y de precisión.

Se siguió el procedimiento general para estimar o aprender un modelo (Gonzalo & De la Cruz García, 2010) a partir de datos. Las fases de formulación del problema e hipótesis iniciales no se detallan ya que en los capítulos anteriores se desarrollaron, las fases que aquí tratan son: Generación/Recopilación de datos, Pre-Procesamiento, Estimación y Validación del modelo.

Generación/recopilación de datos

Dado que no existe un corpus de conversaciones con contenido pedófilo disponible para investigación, fue necesario la construcción del mismo utilizando diferentes colecciones. Es así que uno de los principales aportes de esta investigación es la recopilación del corpus de conversaciones tanto positivas como negativas de conversaciones con contenido pedófilo

Recolección de corpus de conversaciones positivas (con contenido pedófilo)

Para la recolección del corpus de conversaciones positivas se utilizaron 533 conversaciones disponibles en el sitio www.justiced-pervverted.com⁴ relacionadas parcialmente en la tabla 1.

El sitio web www.justiced-pervverted.com es una iniciativa que recluta voluntarios para posar como víctimas (menores de edad) y posteriormente publicar las conversaciones que estos establecen en *chats* con posibles depredadores (pederastas). Existen otros voluntarios que clasifican o califican de 1- 5 dichas conversaciones de acuerdo al nivel contenido de "adulación maliciosa". En algunos casos publican los arrestos realizados por la policía de dichas personas que efectivamente eran pederastas, ya que iban hacer contacto físico con las supuestas víctimas.

La recolección de estas conversaciones fue en forma manual una a una, ya que el sitio no prestaba ningún servicio para la descarga en forma masiva.

Las conversaciones se guardaron en archivos individuales para cada una de ellas y se etiquetaron con el nombre de convicto o depredador (alias o *nickname* de la persona) según el sitio.

Dado que cada conversación tiene asociado un valor promedio 1 a 5 siendo 5 el promedio más alto nivel de "adulación maliciosa" (*Slimyness*, en inglés) según calificación emitida por usuarios voluntarios del sitio web.

Para la construcción del corpus se incluyeron aquellas conversaciones con calificación superior a 3.0, en virtud de que en estas se evidencia contenido pedófilo.

Tabla 1. Relación parcial de archivos de conversaciones positivas

	Nombre Archivo Conversación	Promedio (Slimyness)
1	fleet_captain_jaime_wolfe	4,89
2	thenewperson62	4,88
3	jackman_9682	4,88
4	vamale_692005	4,87
5	kalowoodsman	4,86
6	DavieWants2	4,86
7	smileman74	4,75
8	shinelfmc2005	4,73
9	swgamaleyess	4,72

Fuente: Autores

Debido a que las conversaciones provienen de diferentes fuentes como *Yahoo*, *Google*, *Hotmail* entre otras, aparecen con diferentes formatos y diferentes atributos como fechas, horas, personas, entre otros. En la tabla 2 se muestra el contenido parcial de una de las conversaciones.

⁴ Disponible en www.justiced-pervverted.com

Tabla 2. Ejemplo de Conversación con contenido pedofílico descargada de www.perverted-justice.com

```

1 I checked with the boss, added him to my yahoo friends list, and started chatting. He wasted no time,
2 as you'll soon see.
3
4 05-11-2005
5
6 baby_breana (3:10:35 PM): hi u sent me a email so i added u to my list
7
8 baby_breana (3:55:38 PM): hi
9
10 baby_breana (4:13:00 PM): hi
11
12 aladiesmasseur (4:13:13 PM): hello there my sweet
13
14 baby_breana (4:13:27 PM): is ur name jim?

```

Fuente: información tomada de www.justiced-perverted.com

Recolección de corpus de conversaciones negativas (sin contenido pedófilo)

Para la recolección de la colección de conversaciones negativas se utilizaron conversaciones de dos fuentes:

- El NPS Chat Corpus, Release 1.0 (Forsythand & Martell, 2007) consistente en 10,567 líneas de conversación (*posts*) tomados de 500,000 *posts* que se recolectaron de varios servicios online. Este corpus está según formato XML y compuesto por 16 archivos, en la figura 1 se muestra parcialmente un ejemplo de una conversación.

```

1 <!-- edited with XMLSpy v1007 sp1 (http://www.altova.com) by Eric Forsyth (Naval Postgraduate School) -->
2 <Session xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:noNamespaceSchemaLocation="postClassPOSTagset.xsd">
3   <Posts>
4     <Post class="Statement" user="10-19-20aUser7">now im left with this gay name</terminal>
5       <pos "RB" word="now"/>
6       <pos "PRP" word="im"/>
7       <pos "VBD" word="left"/>
8       <pos "IN" word="with"/>
9       <pos "DT" word="this"/>
10      <pos "JJ" word="gay"/>
11      <pos "NN" word="name"/>
12    </terminal>
13  </Post>
14  <Post class="Emotion" user="10-19-20aUser7">P</terminal>
15    <pos "IN" word="P"/>
16  </terminal>
17 </Post>
18 <Post class="System" user="10-19-20aUser76">PART</terminal>
19   <pos "VB" word="PART"/>
20 </terminal>
21 </Post>
22 <Post class="Greet" user="10-19-20aUser59">hey everyone </terminal>
23   <pos "UH" word="hey"/>
24   <pos "NN" word="everyone"/>
25 </terminal>
26 </Post>
27 <Post class="Statement" user="10-19-20aUser113">ah well</terminal>
28   <pos "UH" word="ah"/>
29   <pos "NN" word="well"/>
30 </terminal>
31 </Post>
32 <Post class="System" user="10-19-20aUser7">NICK :10-19-20aUser7</terminal>
33   <pos "NN" word="NICK"/>
34   <pos "NN" word=""/>

```

Figura 1. Ejemplo de conversación sin contenido pedofílico

Fuente: Corpus NPS Chat (Forsythand & Martell, 2007).

- El corpus ElsnerCharniak08 (Martin et al., 1986) contiene 20,419 posts correspondientes 52:18 horas de conversaciones en un chat grupal sobre Linux. Como pre-procesamiento de este corpus, se separaron en archivos por cada nick utilizado.

Pre-procesamiento de las conversaciones

Como se observó en el apartado anterior los archivos extraídos tienen diferentes estructuras, por lo que es necesario un pre-procesamiento con el fin de normalizarlos y eliminar elementos que no aportan información al proceso de clasificación. Dentro de las actividades del pre-procesamiento se desarrollaron los siguientes procesos:

- **Normalización de los textos de los diferentes chats.** Dado que el corpus se formó a partir de diferentes archivos provenientes de distintas fuentes y con diferentes formatos, fue necesario realizar en primera instancia una tarea

de normalización de texto con el fin que todos los archivos quedaran únicamente con las líneas de conversación. Se hizo necesario realizar una tarea exhaustiva de normalización de textos, de igual forma fue necesario construir expresiones regulares muy particulares para tratar casos particulares. A continuación se resumen los pasos que se aplicaron:

- Eliminación de nombres de usuarios (“nick-names”) de los post o líneas de conversación.
- Eliminación de fecha y hora de las líneas de conversación.
- Eliminación de líneas de anotaciones de los archivos de conversaciones.
- Eliminación de etiquetas XML, en el caso del corpus NPS Chat Corpus, Release 1.0.
- Separación en archivo por cada conversación del corpus ElsnerCharniak08.
- Unificación de la codificación de los archivos al código ANSI (American National Estandar Institute)

Para el desarrollo de estas actividades se desarrollaron programas en java que eliminaron la información irrelevante de los archivos e hicieron las transformaciones.

- **Elaboración de Listado de palabras vacías Stop Words.** Para el desarrollo de esta tarea se elaboró en forma manual un listado de 67 Palabras Vacías (Stop Words) relacionadas con el contexto, ya que para el caso de la presente investigación existen palabras que se repiten en todas las conversaciones y que no aportan valor discriminatorio al clasificador.
 - **Pre-procesamiento de lenguaje Natural.** Una vez ya normalizados los textos de las conversaciones se desarrollaron tareas propias del procesamiento del lenguaje natural, tales como: Eliminación de *stop Word*, *Stemming*, colocar el texto en minúsculas, uso de frecuencias TF/IDF.

Carga y verificación de archivos de conversaciones. Se cargaron 533 conversaciones con contenido pedófilo y 650 conversaciones sin contenido pedófilo, para un total de 1183 conversaciones.

Revisión de las frecuencias de las palabras. A partir de un análisis de las frecuencias se revisó y corrigió el corpus para evitar la existencia de palabras o símbolos que no estuvieran directamente relacionados con las conversaciones como fechas de chats, anotaciones de los agentes, mensajes del sistema y demás información no relevante para el análisis.

Aplicación de tareas de Pre procesamiento. Para el desarrollo de esta tarea se probaron las diferentes opciones de filtros de Weka, a través de la función de pre-procesamiento: *StringToWordVector*, (que convierte los archivos de texto en forma vectorial). En la tabla 3 se detallan los parámetros que se modificaron.

Tabla 3. Parámetros de Pre-procesamiento

<i>IDFTransform</i>	Si está en Verdadero, las frecuencias de las palabras serán transformadas en:
<i>TFTransform</i>	Si está en verdadero, las frecuencias de las palabras serán transformadas en: $\log(1 + \text{frecuencia de la palabra } i \text{ en la instancia } j)$
<i>Stemmer</i>	Se hizo prueba sin <i>stemming</i> y con los siguientes algoritmos: <i>IteratedLovinsStemmer</i> (Lovins, 1968) <i>SnowballStemmer</i> (Chinchor, 1997)
<i>Stop Words</i>	Se hicieron las siguientes variaciones: Sin stopWords StopWord weka StopWord Personalizado

Fuente: Autores

Diseño sistema inteligente

Una vez definido el clasificador, se propone una arquitectura para el sistema inteligente y su posterior implementación en la plataforma JADE (Java Agent Development Framework) (Bellifemine, Poggi, & Rimassa, 2001). En la figura 2 se muestra a través de un diagrama de actividades de UML, un esquema general del sistema.

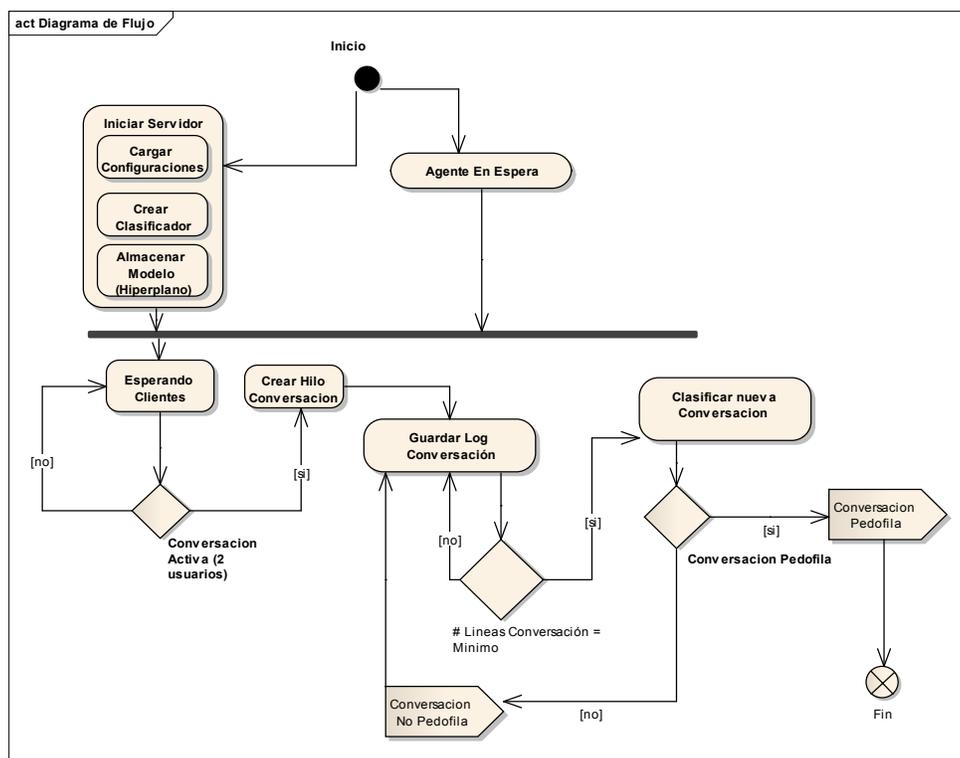


Figura 2. Diseño detallado del sistema

Fuente: Autores

Tal como se detalla en la figura 2, el sistema inicia colocando el servidor chat en línea, dicho servidor lee los archivos de configuración y de entrenamiento, crea el clasificador y queda a la espera de un cliente.

Inmediatamente un cliente se conecta al servidor, se crea un hilo para dicha conversación, este empieza a leer y guardar las líneas de la conversación en un archivo, apenas este archivo supere el número mínimo de líneas especificado en el archivo de configuración, es enviado al clasificador para determinar si es una conversación con contenido pedófilo.

Resultados

Una vez realizada la carga de los archivos de conversaciones y realizado las diferentes tareas de pre-procesamiento se realizaron las diferentes opciones de variación de los parámetros en los algoritmos de clasificación.

Se ejecutaron tres algoritmos (Máquina de Vectores de Soportes, K-NN y *Naive Bayes*) para cada una de las variaciones de los parámetros de pre-procesamiento y se registraron las medidas de efectividad. En la tabla 4 se muestran los resultados, donde se evidencia que con el algoritmo de Máquina de Vectores de soporte se logran los mejores resultados. La variación en el número de la dimensionalidad como resultado de los diferentes

opciones de pre- procesamiento no altera en forma significativa los resultados y si por el contrario el tiempo en que se demoraba la ejecucion en hacer la clasificacion.

Tabla 4. Resultado de los valores de eficiencia para los diferentes algoritmos y parámetros

Algoritmo	IDF	TFT	Stem	stopwords	useStoplist	atributos	TP - Rate	FP - Rate	Precision	Recall	F-Measure
Naive Bayes	false	false	null	weka-3.7	false	1.444	95,000%	5,000%	95,000%	95,000%	95,000%
BayesNet	false	false	null	weka-3.7	false	1.444	95,100%	4,900%	95,100%	95,100%	95,100%
SMO	false	false	null	weka-3.7	false	1.444	98,000%	2,100%	98,000%	98,000%	98,000%
k-NN(K=1)	false	false	null	weka-3.7	false	1.444	92,000%	6,900%	92,800%	92,000%	92,000%
K-NN(K=2)	false	false	null	weka-3.7	false	1.444	88,300%	10,100%	89,800%	88,300%	88,300%
k-NN(K=3)	false	false	null	weka-3.7	false	1.444	85,000%	12,700%	87,800%	85,000%	85,000%
K-NN(K=4)	false	false	null	weka-3.7	false	1.444	82,700%	14,700%	86,200%	82,700%	82,500%
K-NN(K=5)	false	false	null	weka-3.7	false	1.444	82,100%	15,200%	85,900%	82,100%	81,900%
Naive Bayes	false	false	IteratedLovinsStem	weka-3.7	false	1.432	94,900%	5,000%	94,900%	94,900%	94,900%
BayesNet	false	false	IteratedLovinsStem	weka-3.7	false	1.432	95,000%	5,000%	95,000%	95,000%	95,000%
SMO	false	false	IteratedLovinsStem	weka-3.7	false	1.432	97,500%	2,500%	97,500%	97,500%	97,500%
k-NN(K=1)	false	false	IteratedLovinsStem	weka-3.7	false	1.432	90,300%	8,200%	91,700%	90,300%	90,300%
K-NN(K=2)	false	false	IteratedLovinsStem	weka-3.7	false	1.432	88,200%	10,000%	90,100%	88,200%	88,200%
k-NN(K=3)	false	false	IteratedLovinsStem	weka-3.7	false	1.432	83,900%	13,500%	87,400%	83,900%	83,800%
K-NN(K=4)	false	false	IteratedLovinsStem	weka-3.7	false	1.432	82,300%	14,800%	86,500%	82,300%	82,200%
K-NN(K=5)	false	false	IteratedLovinsStem	weka-3.7	false	1.432	81,500%	15,600%	85,800%	81,500%	81,300%
SMO	true	true	IteratedLovinsStem	stopwork.txt	true	39.229	98,100%	2,000%	98,100%	98,100%	98,100%
Naive Bayes	true	true	IteratedLovinsStem	stopwork.txt	true	39.229	94,300%	5,900%	94,300%	94,300%	94,300%
K-NN(K=3)	true	true	IteratedLovinsStem	stopwork.txt	true	39.229	76,200%	20,500%	81,100%	76,200%	75,900%
SMO	true	true	SnowballStemmer	weka-3.7	false	16.599	97,000%	3,000%	97,000%	97,000%	97,000%

Fuente: Autores

En la figura 3 se muestra la curva ROC (*Receiver Operating Characteristic*) relacionando en el eje x la especificidad (Rata de Falsos Positivos) y en el y la sensibilidad (Rata de Verdaderos Positivos), tanto en la clasificación de conversaciones negativas como conversaciones positivas, utilizando el método de Maquinas de Vectores de Soporte (*SVM Support Vector Machine*).

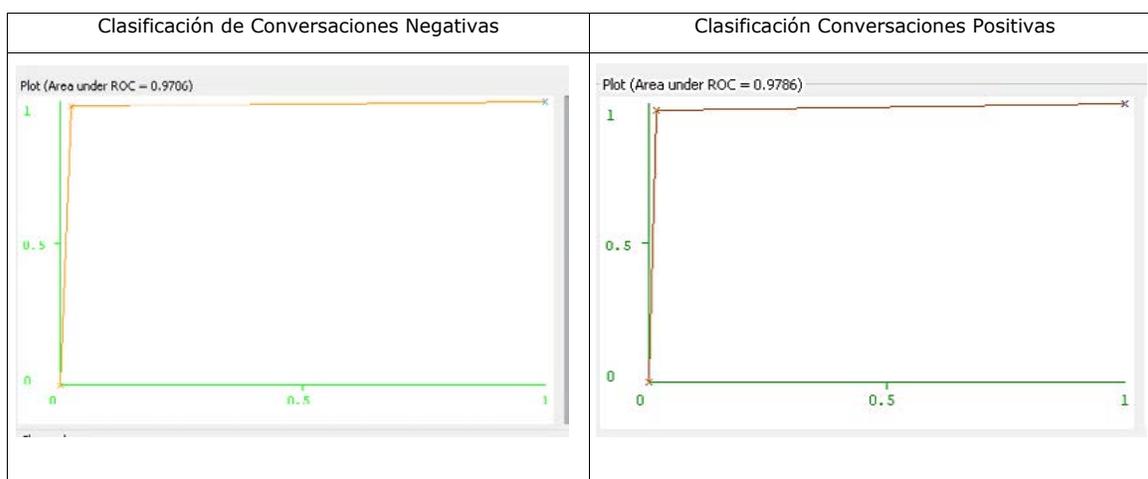


Figura 3. Curva ROC con el método SVM (Support Vector Machine)

Fuente: Autores

Como se observa en la figura 3, el clasificador SVM se comporta en forma similar tanto en el caso de para el caso de las instancias positivas como negativas obteniendo un valor de ROC(0,9786), es decir el 98% de las conversaciones tanto positivas como negativas se clasifican correctamente.

En forma semejante, en la figura 4 se muestra la curva ROC para el caso del método Naive Bayes, el mejor segundo clasificador, en este caso el comportamiento para la clasificación de las instancias positivas (0,9634) es un poco mejor que las negativas (0,9611)

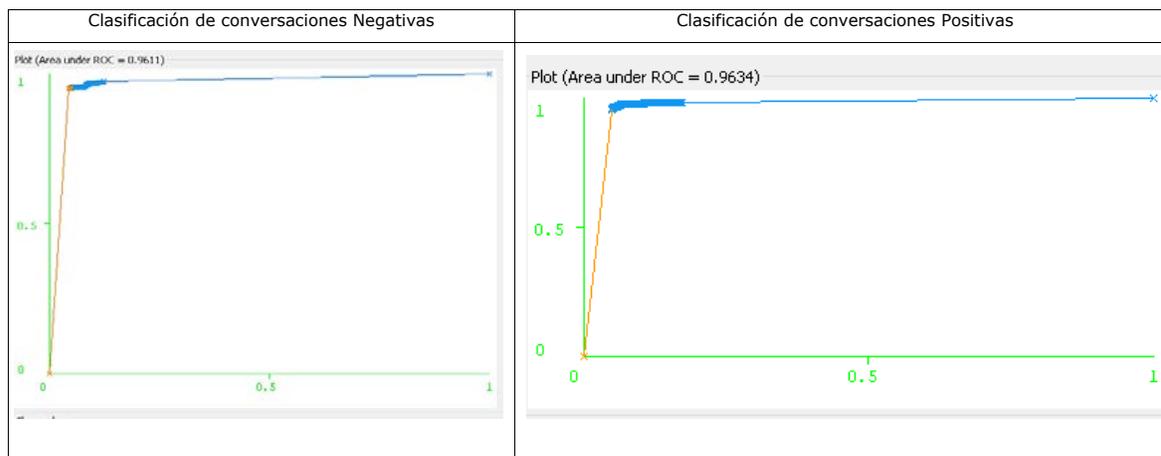


Figura 4. Curva ROC para el método Naive Bayes

Fuente: Autores

Discusión

A partir de los datos obtenidos y tabulados en la tabla 4 correspondientes a los diferentes experimentos desarrollados, se puede evidenciar que acorde con la literatura sobre clasificación de textos, con el método de Maquinas de Vectores de Soporte se logran mejores resultados de eficiencia. El método SMO (*Sequential Minimal Optimization*) utilizado como entrenamiento de SVM (*Support Vector Machine*) muestra un mejor comportamiento de f-measure (98.1%), combinación de Recall y Precision, que los demás métodos, independiente de la aplicación de técnicas para reducción de la dimensionalidad tales como *stemming*, *stop words* y *TD-IDF* que no afectan significativamente la eficiencia. Adicionalmente, la ampliación del número de atributos tampoco tiene impacto significativo en este algoritmo.

Los mejores resultados, en el método de Maquinas de Vectores de Soporte, se logran aplicando el archivo de stopwords creado específicamente para el proyecto, y utilizando el algoritmo de *stemming* Lovins desarrollado por (Lovins, 1968).

Mientras si bien, los métodos de Naive Bayes y k-NN, le sigue en orden de eficiencia (95% y 85% respectivamente) dado por el valor de f-measure, estos son más sensibles a variación de parámetros tales como el número de *k* (en el caso de *k-nn*) y técnicas de reducción de dimensionalidad, mencionadas anteriormente.

El algoritmo de *k-nn* vecinos, presenta una variación considerable con respecto a los valores de *y* a los valores de los parámetros de pre-procesamiento, tales como *stemming*, selección de atributos, observándose que se obtiene un valor de f-measure = 75,9% (el peor resultado) para el caso *k=3* cuando se toman 39.299 atributos y aplicando el *stemming* Lovins, y el mejor resultado f-measure= 85%, en el caso *k= 2*, cuando no se aplica *stemming* y se toman 1.444 atributos

Conclusiones

Se diseñó e implementó un prototipo de agente que permite clasificar en forma automática, partir del texto de una conversación, si dicho contenido es pedófilo o no, acorde a los objetivos planteados al inicio del proyecto.

Se confirmó el algoritmo de Maquinas de Vectores Soporte como el método con mejor precisión para la clasificación de textos con alta dimensionalidad y tolerante a variación léxica de los textos, que se evidencia en el poco impacto que tiene en la precisión cuando se utilizan técnicas de pre procesamiento tales como stemming, o la utilización de stop words que evitan o reducen la variación léxica.

El algoritmo de clasificación con mejores resultados fue el de máquinas de vectores de soporte con una precisión del 98%, aunque otros algoritmos obtuvieron resultados cercanos a este valor, fueron más sensibles a las técnicas de pre-procesamiento.

Se integraron técnicas de Aprendizaje Maquina, Procesamiento del lenguaje Natural y de Agentes software, disciplinas que generalmente se trabajan en forma independiente.

Desde el aspecto social, se propone un prototipo funcional para la protección de niños contra la pedofilia en las salas de chat, que en países como Colombia es un tema poco investigado y en algunos aspectos tabú.

Aunque esta investigación es de carácter científico y tecnológico, toca aspectos sociales como la pedofilia que tiene un fuerte impacto en la vida de los niños que son abusados, ya que estos niños no volverán a ser iguales y los marcará por el resto de sus vidas, y cualquier aporte que se pueda hacer por evitarlo es bueno.

Difundir esta herramienta en las entidades estatales de Colombia encargadas de hacer esta vigilancia, debido a que es conocido que en nuestro país es reconocido como destino sexual y ciudades como Cartagena existe evidencia de este desagradable comercio con los niños y niñas.

Referencias

- Adams, P. H., & Martell, C. H. (2008). *Topic detection and extraction in chat*. Paper presented at the Semantic Computing, 2008 IEEE International Conference on.
- Apuzzo, M. FBI online sex stings winning first convictions. *Electronic Version USA Today*. http://www.usatoday.com/tech/news/2004-01-25-pedo-stings_x.htm
- Association, A. P., & DSM-IV., A. P. A. T. F. o. (2000). *Diagnostic and statistical manual of mental disorders: DSM-IV-TR*: Amer Psychiatric Pub Inc.
- Baumes, J., Goldberg, M., Magdon-Ismael, M., & Wallace, W. A. (2004). Discovering Hidden Groups in Communication Networks (pp. 378-389).
- Bellifemine, F., Poggi, A., & Rimassa, G. (2001). *JADE: a FIPA2000 compliant agent development environment*. Paper presented at the Proceedings of the fifth international conference on Autonomous agents.
- Bengel, J., Gauch, S., Mittur, E., & Vijayaraghavan, R. (2004). *Chattrack: Chat room topic detection using classification*. Paper presented at the In 2nd Symposium on Intelligence and Security Informatics (in review).
- Breeden, B., & Mulholland, J. (2006). *Investigating 'Internet Crimes Against Children' (ICAC) cases in the state of Florida*. Paper presented at the Proceedings of the 2006 ACM symposium on Applied computing, Dijon, France.
- Burger, J., Cardie, C., Chaudhri, V., Gaizauskas, R., Harabagiu, S., Israel, D., . . . Miller, G. (2001). *Issues, tasks and program structures to roadmap research in question & answering (Q&A)*. Paper presented at the Document Understanding Conferences Roadmapping Documents.
- Çamtepe, S. A., Goldberg, M., ismail, M. M., & Krishnamoorthy, M. (2005). *Detecting conversing groups of chatters: a model, algorithms, and tests*. Paper presented at the In Proceedings of the IADIS International Conference on Applied Computing.

- Çamtepe, S. A., Krishnamoorthy, M. S., & Yener, B. (2004). *A Tool for Internet Chatroom Surveillance*. Paper presented at the Proceedings from the second Symposium on Intelligence and security informatics, 2004., Tucson , Arizona.
- Carranza, D. A. O., & Rodríguez, J. C. (1997). *Sistemas inteligentes artificiales y su aplicación en ingeniería*: Universidad Nacional de Colombia, Sede Medellín, Facultad de Minas.
- Copestake, A., & Jones, K. S. (1990). *Natural language interfaces to databases*: Cambridge Univ Press.
- Chen, H.-C., Goldberg, M., & Magdon-Ismael, M. (2004). Identifying Multi-ID Users in Open Forums (pp. 176-186).
- Chinchor, N. (1997, 1997). *MUC-7 named entity task definition*.
- Durkin, K. F. (1997). Misuse of the Internet by Pedophiles: implication for law enforcement and probation practice. *Federal Probation*, 61, 14-18.
- Espino-Duque, G. P. (2010). Criminalidad en cifras: delitos en Colombia, 2009. *Criminalidad*, 52(1), 15-141.
- Fernández, L., Fernández, C., Victoria, M., Garrido, F., & María, J. (2008). Abusos sexuales a menores: estado de la cuestión a nivel nacional e internacional. *ICEV. Revista d'Estudis de la Violencia*(6), 4.
- Forsythand, E. N., & Martell, C. H. (2007). *Lexical and Discourse Analysis of Online Chat Dialog*. Paper presented at the Proc. International Conference on Semantic Computing ICSC 2007.
- Fulda, J. S. (2002). Do Internet stings directed at pedophiles capture offenders or create offenders? And allied questions. *Sexuality & Culture*, 6(4), 73-100.
- Gonzalo, p. M.-s., & de la cruz garcia, j. M. (2010). Aprendizaje automático (pp. 376): Ra-Ma.
- Hernández, E. F. (2008). Presentación plan nacional de prevención y erradicación de la explotación sexual comercial de niños, niñas y adolescentes.
- Kontostathis, A., Edwards, L., & Leatherman, A. (2009). *ChatCoder: Toward the tracking and categorization of internet predators*.
- Lovins, J. B. (1968). *Development of a stemming algorithm*: MIT Information Processing Group, Electronic Systems Laboratory.
- Lunau, K. How Canada took out the pedophiles. *Maclean's*, 120(25), p33.
- MacKean, L. (2004). Fears Paedophiles will Revel in Cyberspace. *The 7: 30 Report*, 13, 1024959.
- Martin, P., Appelt, D. E., Grosz, B. J., & Pereira, F. (1986). *TEAM: an experimental transportable natural-language interface*. Paper presented at the Proceedings of 1986 ACM Fall joint computer conference.
- Mutton, P. (2004). PieSpy. from <http://www.jibble.org/piespy/>
- Neil W. Van Dyke, H. L., Pattie Maes. *Butterfly: A Conversation-Finding Agent for Internet Relay Chat Irvy*.
- Patterson, P. F. a. A. Paedophile Internet Activity. *AUSTRALIAN INSTITUTE OF CRIMINOLOGY*.
- Pendar, N. (2007). *Toward Spotting the Pedophile Telling victim from predator in text chats*. Paper presented at the Proc. International Conference on Semantic Computing ICSC 2007.
- Penna, L., Clark, A., & Mohay, G. (2005). *Challenges of automating the detection of paedophile activity on the Internet*. Paper presented at the Proc. First International Workshop on Systematic Approaches to Digital Forensic Engineering.
- Romaniuk, S. G. (2000). Using intelligent agents to identify missing and exploited children. *IEEE Intelligent Systems and Their Applications*, 15(2), 27--30.
- Russell, S. J., Norvig, P., Canny, J. F., Malik, J. M., & Edwards, D. D. (1995). *Artificial intelligence: a modern approach* (Vol. 2): Prentice hall Englewood Cliffs.
- Stone, B., & Miller, M. (1999). A HIGH-TECHNOLOGY CRASH. *Newsweek*, 134(14), 54-55.
- Stratica, N., Kosseim, L., & Desai, B. C. (2005). Using semantic templates for a natural language interface to the CINDI virtual library. *Data &*

Knowledge Engineering, 55(1), 4-19.

Unicef. (2010). *UNICEF Informe Anual 2009*: Unicef.

Wang, Y. C., Joshi, M., Cohen, W., & Rosé, C. (2008). *Recovering implicit thread structure in newsgroup style conversations*.

Zitzen, M., & Stein, D. (2004). Chat and conversation: a case of transmedial stability? *Linguistics*, 42(5), 983-1021.