novática
Revista de la Asociación de Técnicos de Informática

Novática, revista fundada en 1975 y decana de la prensa informática española, es el órgano oficial de expresión y formación continua de **ATI** (Asociación de Técnicos de Informática), organización que edita también la revista **REICIS** (Revista Española de Innovación, Calidad e Ingeniería del Software).

<http://www.ati.es/novatica/>
<http://www.ati.es/reicis/>

ATI es miembro fundador de **CEPIS** *(Council of European Professional Informatics Societies)* y es representante de España en **IFIP** *(International Federation for Information Processing);* tiene un acuerdo de colaboración con **ACM** *(Association for Computing Machinery),* así como acuerdos de vinculación o colaboración con **AdaSpain**, **AI2**, **ASTIC**, **RITSI** e **Hispalinux**, junto a la que participa en **ProInnova**.

Coordinación Editorial, Redacción Central y Redacción ATI Madrid
Padilla 66, 3º, dcha., 28006 Madrid
Tlfn.914029391; fax.913093685 <novatica@ati.es>
Composición, Edición y Redacción ATI Valencia
Av. del Reino de Valencia 23, 46005 Valencia
Tlfn. 963740173 <novatica_prod@ati.es>
Administración y Redacción ATI Cataluña
Via Laietana 46, ppal. 1ª, 08003 Barcelona
Tlfn.934125235; fax 934127713 <secregen@ati.es>
Redacción ATI Aragón
Lagasca 9, 3-B, 50006 Zaragoza.
Tlfn./fax 976235181 <secreara@ati.es>
Redacción ATI Andalucía <secreand@ati.es>
Redacción ATI Galicia<secregal@ati.es>
Suscripción y Ventas <novatica.subscripciones@atinet.es>
Publicidad Padilla 66, 3º, dcha., 28006 Madrid.
Tlnf.914029391; fax.913093685 <novatica@ati.es>
Imprenta: Derra S.A., Juan de Austria 66, 08005 Barcelona.
Depósito legal: B 15.154-1975 -- ISSN: 0211-2124; CODEN NOVAEC
Portada: Puente en el cielo - Concha Arias Pérez / © ATI
Diseño: Fernando Agresta / © ATI 2003

sumario

Nº 219, septiembre-octubre 2012, año XXXVIII

Tema del próximo número: "Experiencias y retos en IPv6"

Avances tecnológicos en la protección del menor en redes sociales

José María Gómez Hidalgo[1], Andrés Alfonso Caurcel Díaz[2]
[1]Director de I+D en Optenet, Coordinador de la sección técnica "Acceso y recuperación de información" de Novática; [2]Grupo de investigación de Sistemas Inteligentes para la Movilidad y Comunicación Accesible de la Universidad Politécnica de Madrid

<jgomez@optenet.com>, <caurcel@gmail.com>

Resumen: *En este artículo se presentan los principales riesgos y amenazas a la seguridad y la privacidad de los menores en las redes sociales. Estos riesgos incluyen la exposición a materiales inapropiados, la pornografía infantil y la pedofilia, el acoso escolar, el sexting y la sextorsión, el acoso sexual y la ciber-adicción, entre otros. Asimismo, se describen las iniciativas españolas y europeas tomadas para la auto-regulación de los operadores y proveedores de las redes sociales, y las escasas medidas proactivas tomadas por los mismos para abordar estos problemas. Por último, se revisan algunas iniciativas tecnológicas punteras que permiten detectar situaciones peligrosas como el acoso sexual y el escolar, o las redes de pedófilos que difunden pornografía infantil.*

Palabras clave: *Acoso sexual, bullying, ciber acoso, control parental, grooming, pedofilia, pornografía, protección del menor.*

Autores

José María Gómez Hidalgo es Doctor en Matemáticas por la Universidad Complutense de Madrid, y ha sido profesor y Director del Departamento de Sistemas Informáticos en diversas entidades universitarias durante los últimos 15 años. Es Director I+D en Optenet. Ha participado en más de 20 proyectos de investigación españoles y europeos, varios de ellos como Investigador Principal. Además, ha sido co-autor de una serie de trabajos de investigación y revisor para la Comisión Europea. Sus intereses investigadores se centran en la Minería de Texto aplicada en campos como la información personalizada, la biomedicina, y especialmente, la seguridad informática y la protección del menor en Internet. Es coordinador de la sección técnica "Acceso y recuperación de información" de ***Novática***.

Andrés Alfonso Caurcel Díaz es estudiante de Informática y becario de la Universidad Politécnica de Madrid, dentro del grupo de investigación de Sistemas Inteligentes para la Movilidad y Comunicación Accesible, dirigido por el Dr. José Gabriel Zato Recellado. Andrés realiza labores de investigación en diversos proyectos centrados en la protección del menor en Internet, el transporte inteligente y seguridad del viandante. Sus intereses de investigación abarcan las técnicas de Inteligencia Artificial aplicadas a todos estos temas.

1. Introducción

Hoy en día, los menores de 15 años son considerados "nativos digitales", personas que no conciben el mundo sin Internet. Se ha producido un salto generacional a nivel tecnológico, una suerte de brecha digital, y los niños y jóvenes aprenden cada día a sacar partido del vasto entorno de Internet, y en especial de las redes sociales, para relacionarse, comunicarse, aprender y divertirse. Su modo de afrontar este medio deja de lado la privacidad propia y ajena, y las cautelas propias de la madurez, para exponerlos a graves peligros cuya incidencia es en algunos casos marginal, pero siempre creciente, como el ciberacoso por parte de adultos o de otros menores.

Para afrontar estos peligros siempre presentes en las redes sociales, los operadores de las mismas han adoptado una serie de medidas auto-regulatorias en los últimos años, sobre todo como consecuencia de la presión de las instituciones europeas y con el fin de proteger sus marcas comerciales. Estas medidas se limitan generalmente a la acción bajo denuncia. Por ejemplo, está prohibido que un menor de 14 años tenga cuenta en una red social sin autorización paterna, pero si el niño se da de alta con una fecha de nacimiento de, por ejemplo, 1 de enero de 1900, el operador de la red social no suele hacer absolutamente nada hasta que alguien le comunica que esa cuenta corresponde potencialmente a un menor de 14 años.

Obviamente, tomar un enfoque proactivo (por ejemplo, revisar todas las cuentas sospechosas) exige invertir unos recursos substanciales, sobre todo en términos de personal experto. Para reducir este coste y facilitar la pro-actividad de las redes sociales en la protección del menor, es necesario desarrollar sistemas técnicamente avanzados que permitan detectar los riesgos o problemas de manera semi-automática, obligando por tanto al personal de revisión a examinar solamente un pequeño conjunto de casos potencialmente peligrosos, en lugar de la totalidad de los contenidos de la red social.

En este artículo presentamos algunos de los avances técnicos más recientes en esta línea, algunos de los cuales pueden llegar a servir de ayuda a los operadores de redes sociales en un plazo relativamente corto.

2. Peligros para los menores en las redes sociales

Internet es un medio de uso masivo y que fomenta el anonimato, y en consecuencia, es propenso al abuso. En lo que afecta específicamente a los menores, es importante tener en cuenta que numerosos peligros no están vinculados con Internet exclusivamente, sino que se producen en primer lugar en el mundo real y luego pasan al virtual (por ejemplo, el acoso entre iguales), o se producen en ambos entornos indistintamente, y en ocasiones de manera entremezclada (como en el caso del acoso sexual).

De acuerdo a varios estudios ([1][2][3][4]), los menores se enfrentan a los siguientes peligros en Internet en general: Generación y exposición a material inapropiado o ilegal (pornografía, promoción de la violencia, sectas, promoción de la anorexia y la bulimia, etc.); contactos y relaciones de índole sexual con adultos (acoso sexual, o "*grooming*"); acoso por parte de iguales (acoso escolar o "*bullying*"); *sexting* y *sextorsión* (intercambio de materiales gráficos propios sensibles); difusión de pornografía infantil; problemas de adicción; abusos de la privacidad; y acceso y uso de aplicaciones comerciales ilegales para su edad (e.g. juegos de casino).

Es obvio, y las estadísticas así lo confirman, que los menores y adolescentes son usuarios intensivos de las redes sociales. Por ejemplo, en EE.UU. los estudios permiten contrastar el uso de redes sociales por parte de adolescentes (12-17 años) y jóvenes adultos (18-29 años), situado en el 72-73% de los encuestados, con el realizado por parte de los adultos (30+ años), situado en el 40% [5]. En Europa, el 38% de los niños (9-12 años) y el 77% de los adolescentes (13-16 años) tienen un perfil en alguna red social [6]. En concreto, en España el 56% de los menores

(9-16 años) afirma tener un perfil propio en una red social [7].

Lógicamente, cabe hacerse la pregunta de si este hecho afecta a los peligros anteriormente mencionados, que están identificados en Internet *en general*. La intuición así lo indica, ya que las redes sociales son entornos de socialización, de generación de contactos y comunicación con ellos, y de compartición de contenidos. Es decir, cabe pensar que:

- Los menores pueden establecer contacto con extraños con más facilidad, lo que puede aumentar la incidencia de fenómenos como el acoso sexual.
- Se pueden compartir contenidos inapropiados con mucha más facilidad, lo que puede llevar a una mayor exposición de los menores a estos contenidos.

Estas intuiciones se ven confirmadas en parte. Por ejemplo, a nivel europeo, uno de cada cuatro niños tiene contacto en Internet con alguien que está fuera de su círculo social (familia, vecinos, amigos). Sin embargo, muy pocos se han visto en el mundo real con alguien de una edad superior a la suya, y apenas un 1% ha tenido una experiencia negativa en este sentido [6]. Entre los menores españoles usuarios de Intenet, un 9% han tenido contacto *offline* con alguien que han conocido *online* [7]. No parece, pues, que a pesar de las continuas noticias en los medios de comunicación sobre abusos por parte de adultos que se hacen pasar por menores en redes sociales, este hecho se produzca con excesiva frecuencia.

En cuanto a los contenidos generados por los usuarios, cabe reseñar que los menores no generan contenido sino que lo comparten [6]. Sin embargo, no parece que se compartan o se accedan a más contenidos inapropiados, sino casi al contrario, ya que por ejemplo, el 19% de los menores afirman haber accedido a contenidos potencialmente perjudiciales generados por otros usuarios (incitación al odio, pro-anorexia, drogas, etc.). Aún menos frecuente es el acceso a imágenes sexuales (11%) [7].

Aunque no existen estadísticas concretas, sí que existen evidencias de que las redes sociales están siendo usadas por pederastas y pedófilos para intercambiar pornografía infantil. No parece sin embargo que el uso de las redes sociales por parte de estas personas haga aumentar la incidencia de este fenómeno[1].

Por otra parte, lo que parece claro es que las limitaciones legales existentes con respecto al acceso a redes sociales por parte de menores de 14 años no tienen resultado. El mismo informe mencionado anteriormente, que indica que a nivel europeo un 38% de los niños entre 9 y 12 años tienen un perfil en alguna red social, así lo evidencia [6].

3. Medidas legislativas y auto-regulatorias de las redes sociales

A la vista de los peligros anteriores, no es de extrañar que exista una presión política y social sobre los operadores de redes sociales, para que ayuden a controlarlos y a proteger al menor en sus sistemas. La presión política se traduce en medidas de carácter regulatorio[2] o legislativo, que tienen un alcance muy limitado principalmente por dos razones:

- La falta de conocimiento técnico para la definición de delitos, o las imprecisiones, vaguedades o limitaciones en los mismos.
- La tras-nacionalidad y falta de acuerdo internacional en los delitos.

Como resultado también de la presión política y social, y especialmente a nivel europeo, se declararon en 2009 un conjunto de "Principios para unas Redes Sociales Más Seguras en la Unión Europea"[3], que fueron firmados por más de 20 operadores de redes sociales (incluyendo Facebook, Google, el operador de Habbo Hotel, Tuenti, etc.) con presencia en Europa. Se trata de los siguientes:

1) Aumentar la concienciación a través de mensajes y de información de carácter educativo para todos los públicos.
2) Trabajar para que los servicios ofrecidos sean apropiados para cada edad, es decir, mostrar qué servicios no son apropiados para determinadas edades y cuáles son los límites de edad establecidos, garantizar que se cumplen los límites de edad tanto legales como del servicio, y promover el uso de controles parentales y de etiquetado de contenidos para adultos.
3) Dotar a los usuarios de herramientas y tecnología para controlar los contenidos relacionados con ellos (e.g. los contenidos puestos por terceros en su "muro").
4) Proporcionar métodos sencillos para reportar conductas o contenidos inapropiados según los Términos de Uso del Servicio.
5) Responder con premura a las notificaciones de contenidos o conductas ilegales.
6) Habilitar y fomentar que los usuarios controlen su privacidad e información personal.
7) Evaluar sus servicios propios de control de contenidos y conductas ilegales o inapropiadas de acuerdo con los Términos de Uso del Servicio.

Se puede observar que, aunque estos principios son un paso adecuado en la dirección de proteger a los menores de los peligros existentes para ellos en las redes sociales, no existe ninguna medida proactiva en la línea de detectar la existencia de los peligros, sino que estas medidas son puramente reactivas. Por ejemplo, el último principio establece que se evaluarán los métodos de control de contenidos y conductas, pero no se especifica cuáles deben ser, al menos de manera mínima, excepto por lo especificado en principios anteriores (actuación bajo demanda).

Por otra parte, además de establecer estos principios, se ha iniciado un programa de vigilancia de los mismos. De hecho, ya se han realizado dos evaluaciones de cumplimiento de los mismos por parte de los firmantes (junio de 2011[4] y septiembre de 2011[5]). Por ahora podemos decir que, aunque los principios han sido firmados, no se cumplen en su totalidad. Obviamente, algunos de ellos requieren un esfuerzo explícito y conllevan un tiempo de implantación, factor que debe ser tenido en cuenta en la evaluación de su cumplimiento.

4. Revisión de las técnicas de análisis más recientes

4.1. Ámbito de la revisión

En la definición de principios anteriores, y en particular en el último de los mismos, se sugieren de manera explícita diversas maneras de controlar el contenido existente en las redes sociales:

- Moderación manual o automática de los contenidos.
- Herramientas técnicas (e.g. filtros) que alerten de contenidos ilegales o prohibidos.
- Alertas generadas por la comunidad.
- Informes producidos por usuarios.

Se puede observar que las dos primeras requieren de la existencia de técnicas eficaces para detectar los contenidos ilegales o inapropiados, ya sea durante su subida a la red social, o cuando sean accedidos por parte de un usuario.

Recientemente (mayo de 2012) se ha hecho pública una nueva estrategia en la agenda digital europea, denominada: "Agenda Digital: Nueva estrategia para mejorar la seguridad en Internet y crear contenidos más adecuados para niños y adolescentes"[6]. En esta estrategia se plantean cuatro objetivos fundamentales:

- Fomentar la producción de contenidos creativos y educativos en línea, destinados a los niños, y desarrollar plataformas que ofrezcan acceso a contenidos adaptados a la edad.
- Aumentar la sensibilización y la enseñanza de la seguridad en línea en todas las escuelas de la UE, a fin de desarrollar la alfabetización digital y mediática de los niños y la autoresponsabilización en línea.
- Crear un entorno seguro para los niños en el que tanto los padres como los menores dispongan de las herramientas necesarias para garantizar su protección en línea (por ejemplo, mecanismos fáciles de utilizar

para denunciar los contenidos y conductas nocivos en línea, parámetros de privacidad predefinidos, adaptados a la edad, que sean transparentes, o controles parentales de fácil utilización).
- Luchar contra el material pornográfico infantil en línea, fomentando la investigación y la utilización de soluciones técnicas innovadoras en las investigaciones policiales.

Nuevamente, los dos últimos objetivos requieren de la existencia o plantean la necesidad de técnicas eficaces para la detección de contenidos ilegales o inapropiados.

Es por ello que presentamos aquí los últimos avances técnicos para analizar los contenidos presentes en las redes sociales. En particular, nos centramos en los siguientes aspectos:

- Verificación de la edad de los usuarios de una red social.
- Detección de contenidos relacionados con el acoso sexual.
- Detección de contenidos relacionados con el acoso escolar.
- Detección de contenidos de pornografía infantil.

Dejamos fuera de este análisis los sistemas de filtrado o de detección de contenidos inapropiados en general (pornografía, pro-anorexia, etc.), porque estos se tratan exhaustivamente en artículos previos [8][9]. Cabe reseñar a este respecto que las técnicas utilizadas en los filtros de contenidos inapropiados son básicamente la utilización de bases de datos de URLs mantenidas manualmente y organizadas por categorías, y la utilización de clasificadores de texto basados en algoritmos de aprendizaje. Esta segunda tecnología, con algunos matices, es la utilizada en muchos de los trabajos mencionados más abajo para la detección de los peligros anteriores.

4.2. Técnicas de detección de la edad

Hoy por hoy, la detección de la edad de una persona a través de un dispositivo electrónico es aún un problema abierto, y que atrae bastante atención por parte de los expertos en biometría. Recordemos que la biometría pretende identificar de manera unívoca a una determinada persona en función de atributos físicos o de su comportamiento.

Las técnicas de análisis biométrico pueden igualmente aplicarse no ya a identificar una persona concreta, sino a aproximar la edad de una persona cualquiera. De hecho, existen diversos trabajos que exploran el uso de rasgos faciales [10][11][12][13][14][15], voz [16], forma de caminar [17], etc.

La cuestión es cuáles de estas técnicas pueden usarse para determinar la edad de una persona en una red social. Teniendo en cuenta que la tecnología biométrica actual puede permitir la identificación muy precisa del usuario de un sistema, y la estimación aproximada de la edad del mismo, se pueden plantear los siguientes casos de uso en lo que respecta al acceso a redes sociales:

1) Verificación de la edad del usuario en el momento del alta o registro.
2) Verificación de la identidad del usuario en el momento del acceso al sistema, equivalente a la verificación continua o periódica de la identidad del usuario.
3) Verificación continua o periódica de la edad del usuario.

El primer caso presenta los siguientes problemas:

- Se exige la disponibilidad de un sensor de toma de datos durante el alta en la red social. El dispositivo de registro (un PC, un móvil) puede no estar dotado de un sensor adecuado.
- En el caso de los *smartphones*, que sí poseen un sensor (e.g. cámara), puede carecer de la precisión necesaria para que los datos sean fiables.
- Se exige al usuario la entrega de información confidencial (e.g. una foto de la cara).

En resumen, se exige al usuario la disponibilidad de un dispositivo con un sensor adecuado (lo que no siempre es posible y limitaría el acceso al servicio), y la entrega de información confidencial y privada. Este caso de uso supone por tanto un limitador en el acceso al servicio, contribuyendo a la brecha digital y planteando problemas de accesibilidad, siendo asimismo un atentado a la privacidad del usuario. Como colofón, no existen garantías de que el sistema sea preciso.

En el segundo caso, se presentan los mismos problemas que en el primero, ya que es preciso realizar una toma de datos de identidad durante el registro, a fin de verificar la misma ocasionalmente y siempre que el dispositivo de acceso posibilite la toma de datos en el momento del acceso al sistema o durante la operación del usuario en la red social.

Sin embargo, el tercer caso es mucho más favorable que los anteriores, ya que:

- No se requiere el uso de dispositivos con capacidades específicas, tan solo aquellos que ya esté utilizando el usuario de la red social.
- No se invade la privacidad del usuario, ya que sólo se utilizan aquellos datos que él mismo ha hecho accesibles en la red social.
- Existen perspectivas favorables para la estimación precisa de la edad del usuario por cuanto la cantidad de datos disponibles sobre el mismo es grande y se extiende en el tiempo.
- Existe la posibilidad de adquirir datos de los usuarios de manera masiva, acrecentando la probabilidad de que los métodos basados en la existencia de colecciones de datos de entrenamiento sean muy efectivos.

Por tanto, son del máximo interés aquellos métodos que permitan tratar de determinar la edad de un usuario en función de la información que comparte voluntariamente en una red social. Aunque los usuarios comparten muy diversos tipos de información en las redes sociales, parece bastante claro que las principales fuentes de información a tratar son el texto y la imagen, ya en lo que se refiere a video y audio, los usuarios suelen compartir contenidos de terceros más que propios.

4.2.1. Detección basada en análisis textual

En cuanto al procesamiento de texto, existen varios trabajos recientes muy relevantes [18[[19]. En casi todos ellos se aplican técnicas de aprendizaje automático sobre diversas representaciones de los textos utilizados. El objetivo es, a partir de una serie de textos de usuarios de distintas edades, construir modelos del tipo de lenguaje característico de cada edad, en función del contenido de los textos (palabras y expresiones usadas) y de los atributos lingüísticos propios de cada edad (uso de emoticonos y expresiones tipo SMS, mayúsculas y abreviaturas, complejidad de las construcciones sintácticas, etc.). Casi todos los trabajos se han realizado sobre el idioma inglés.

Uno de los primeros trabajos es [19], donde se presentan diversos experimentos sobre una colección de datos denominada *NPS Chat Corpus*, *Release* 1.0 , que actualmente es pública. Esta colección consiste en una selección de 10.567 *posts* obtenidos en diversos servicios *online* de chat, formateados en XML, y enmascarados manualmente para preservar la privacidad de los usuarios. Los *posts* han sido etiquetados sintácticamente (nombres, adjetivos, etc.), y también de acuerdo al acto dialogal (despedida, pregunta, emoción, etc.). Todos los textos incluidos en esta colección provienen de chats específicos de la edad, es decir, orientados a edades específicas. En la colección se incluyen *posts* de usuarios adolescentes (archives "*teens*"), y adultos de distintas edades (20-29, 30-39, etc.). Cada archivo de la colección es un registro de las conversaciones de una sala de charla durante un corto periodo de un día determinado. El nombre del archivo incluye esta información junto con el número de *posts* incluidos en dicho archivo. En la **figura 1** se muestra unos ejemplos de *posts* incluidos en la colección.

De entre los enfoques evaluados en [19], el que mejores resultados obtiene es usar una representación del texto que emplea frecuencias de uso de secuencias de tres palabras o trigramas, y *Support Vector Machines (SVM)*

```
<Post class="Statement" user="10-19-20sUser115">
He drew a girl with legs spread<terminals>
   <t pos="PRP" word="he"/>
   <t pos="VBD" word="drew"/>
   <t pos="DT" word="a"/>
   <t pos="NN" word="girl"/>
   <t pos="IN" word="with"/>
   <t pos="NNS" word="legs"/>
   <t pos="VB" word="spread"/>
   </terminals>
</Post>
<Post class="Statement" user="10-19-20sUser7">
hope he didnt draw a penis<terminals>
   <t pos="VBP" word="hope"/>
   <t pos="PRP" word="he"/>
   <t pos="VBD" word="didnt"/>
   <t pos="VB" word="draw"/>
   <t pos="DT" word="a"/>
   <t pos="NN" word="penis"/>
   </terminals>
</Post>
```

Figura 1. Ejemplos de fotografías en la base de datos FG-NET.

como algoritmo de aprendizaje. Los resultados alcanzados son de un 99,6% de efectividad usando la métrica F_1, que combina otras dos técnicas habituales en Recuperación de Información, que son la cobertura (*recall*) y la precisión (*precision*).

En [18] se revisa una serie de trabajos anteriores de otros autores, además de presentar experimentos sobre chats extraídos de la red social belga Netlog. En estos experimentos, la representación de los textos es más adecuada al idioma holandés, y utiliza no sólo desde unigramas a trigamas de palabras, sino también desde unigramas hasta tetragramas de letras. Las clases de edades que se definen no son afines a las salas de chat orientadas a la edad como en [19], sino otras orientadas a la detección de acoso sexual como en trabajos que revisaremos más adelante. También se considera el género de los interlocutores, por lo que se plantean problemas de clasificación como "niñas menores de 16 años vs. hombres mayores de 25 años". Utilizando también SVM, los autores logran sus mejores resultados con unigramas (palabras aisladas) para la tarea de "menores de 16 años vs. mayores de 25 años", con una precisión del 88,2%.

Las métricas usadas en ambos trabajos no son directamente comparables, pero lo que parece claro es que representaciones relativamente sencillas basadas en secuencias de palabras, pueden dar resultados bastante buenos y robustos para los idiomas occidentales.

Recientemente, en el proyecto WENDY[7], se ha desarrollado un clasificador de texto que combina desde unigramas a trigramas de palabras con el uso de información de lenguaje SMS y lingüística en la representación del texto, y SVM como algoritmo de aprendizaje. Se han realizado experimentos en español usando no sólo comentarios, sino también información de preferencias sobre música, películas, libros y *hobbies* de 2.784 usuarios de la red social Tuenti. Estos usuarios se han dividido en rangos de edades de menos de catorce años (75 usuarios), mayores de edad (884 usuarios), y mayores de catorce años y menores de edad (1.825 usuarios). Los resultados obtenidos son muy buenos, ya que se ha logrado alcanzar una F_1 de 100% para los usuarios menores de 14 años, que son los más sensibles de acuerdo con la normativa actual.

4.2.2. Detección basada en análisis facial

Dentro de las redes sociales, las imágenes de caras son muy habituales. Es significativo que los usuarios usen fotos propias con mucha frecuencia como foto del perfil o de su avatar, y que dichas fotos sean en solitario y en primer plano. Además, existe la posibilidad de etiquetar a terceros en las fotos propias, y de hecho Facebook ha comprado Face.com, una empresa de análisis de imagen y reconocimiento facial, con el fin de sacar máximo partido de esta información.

Por tanto, tiene mucho sentido aplicar técnicas de reconocimiento facial e identificación personal a las fotografías disponibles en las redes sociales. Sin embargo, escapa al objetivo del artículo revisar todo el estado de la técnica de la biometría basada en reconocimiento facial. En primer lugar, resulta relevante confirmar la hipótesis de que es posible estimar la edad de una persona usando como dato de entrada la fotografía de su cara. En [10] se ha realizado un experimento manual usando encuestas a usuarios, en las que se les ha solicitado realizar esta estimación. El experimento se ha realizado con personas occidentales y orientales, con el objeto de apreciar el efecto de factores socioculturales. Los resultados son positivos en lo que se refiere a que la aproximación de edad realizada por humanos no es mala, pero se aprecia una tendencia relevante a su infraestimación, y evidentemente, las estimaciones interraciales son menos precisas.

La mayoría de los trabajos de estimación de la edad basada en imagen pueden clasificarse en tres grupos principales:

- Aquellos que pretenden construir un modelo antropomórfico del envejecimiento [14]. En estos trabajos se utiliza la teoría de desarrollo cráneo-encefálico y el análisis de arrugas faciales para construir un modelo relacionado con el crecimiento, para clasificar la edad del sujeto en uno de varios grupos de edad.
- Aquellos que utilizan un sub-espacio de patrones de envejecimiento (*aging pattern subspace*, término acuñado en [12]). En este

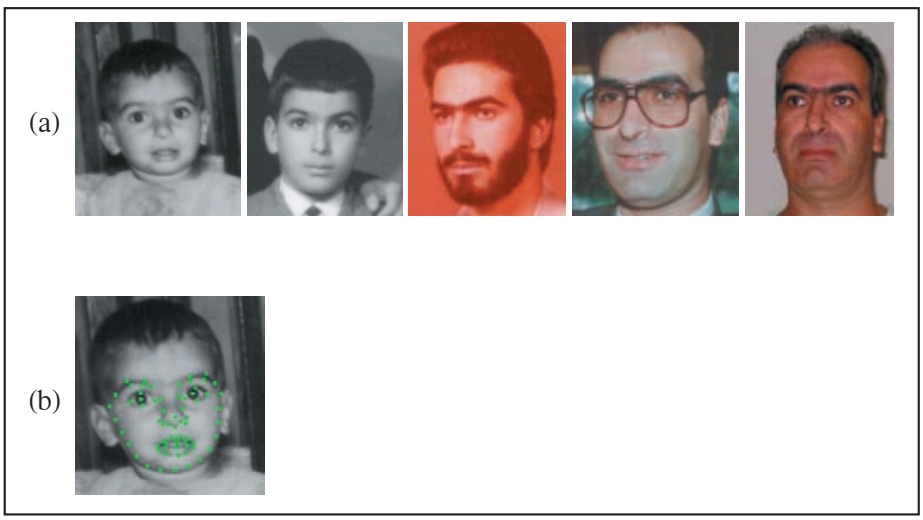

Figura 2. Ejemplos de fotografías en la base de datos FG-NET.

método se definen una serie de parámetros que se ajustan por medio de entrenamiento sobre fotografías de individuos a lo largo de su vida.
■ Aquellos que plantean la estimación de edad como un problema de regresión [11] [13]. En estos métodos se extraen una serie de atributos relativos a la forma y apariencia de la cara, y se construye un modelo con regresión para aproximar la edad de un sujeto objetivo. En la regresión no se suelen utilizar funciones lineales sino estimadores cuadráticos.

En estos trabajos se han obtenido con frecuencia tasas de error de en torno a 5 años en las estimaciones de edad, y los resultados dependen de la cantidad de información disponible en el entrenamiento.

Con el fin de facilitar el trabajo en esta área, y de disponer de una colección de datos que permita la comparación de diversas técnicas, se han creado varias bases de datos de investigación. Una de las más populares es FG-NET[8], que consta de 1.002 imágenes de un total de 82 sujetos diferentes de origen caucásico, tomadas más o menos de frente, y con distintas condiciones de iluminación y color. Una parte significativa de las fotos incluidas está manualmente etiquetada con una serie de puntos de interés que permiten utilizar la forma de la cara como atributos para la utilización de algoritmos de aprendizaje automático. En la **figura 2** se muestran algunas imágenes representativas de la base de datos, formando una sucesión de edad para un mismo sujeto (a), junto con un ejemplo en el que se destacan los puntos de interés (b).

Es obvio que en un entorno como una red social, donde existen franjas de edad claramente diferenciadas de acuerdo a las normativas (menores de 14 años, menores de edad, mayores de edad), un grado de error en la aproximación en torno a 5 años no puede considerarse útil. Igualmente, aunque por ejemplo en las fotos del perfil se pueda esperar encontrar caras con cierta frecuencia, las posiciones y condiciones de iluminación no son precisamente óptimas, y por ejemplo, no permiten detectar los puntos de interés con facilidad. En la **figura 3** se muestra una serie de ejemplos tomados de Tuenti en el proyecto WENDY, donde se puede observar, para distintas edades, ejemplos de fotos de avatares. No se muestran dibujos, fotos de famosos, logotipos, textos (e.g. grafiti), etc., que con frecuencia son usados como fotos de perfil.

En este sentido, en el proyecto WENDY se plantea un enfoque multimodal, es decir, que combine distintas fuentes de información (textos, fotos, etc.) para obtener un clasificador más preciso en las tres clases objetivo, que se corresponden con las franjas de edad anteriores. Recientemente se ha presentado un trabajo realizado en el marco de dicho proyecto [15], en el que se realizan una serie de pruebas iniciales sobre las imágenes de FG-NET, como línea base. En estas pruebas no se utiliza la información de puntos de interés, sino que se caracterizan las imágenes en función directa de sus píxeles, y se realizan distintas pruebas usando esta representación o su reducción con análisis de componentes principales (*Principal Component Analysis*, PCA). Sobre estas representaciones, se entrenan *Support Vector Machines*.

Como los resultados de los algoritmos de aprendizaje dependen claramente de la cantidad de fotos utilizadas en el entrenamiento, los resultados obtenidos para las tres franjas de edad son relativamente pobres, ya que la clase de los menores de 14 años está muy infra-representada. Sin embargo, los resultados de clasificación cuando se plantea el problema de "menores o iguales a 18 años vs. mayores de 18 años", alcanza valores de F_1 del 80% de efectividad. Estos resultados, combinados con los anteriormente mencionados sobre texto, son sumamente positivos y esperanzadores.

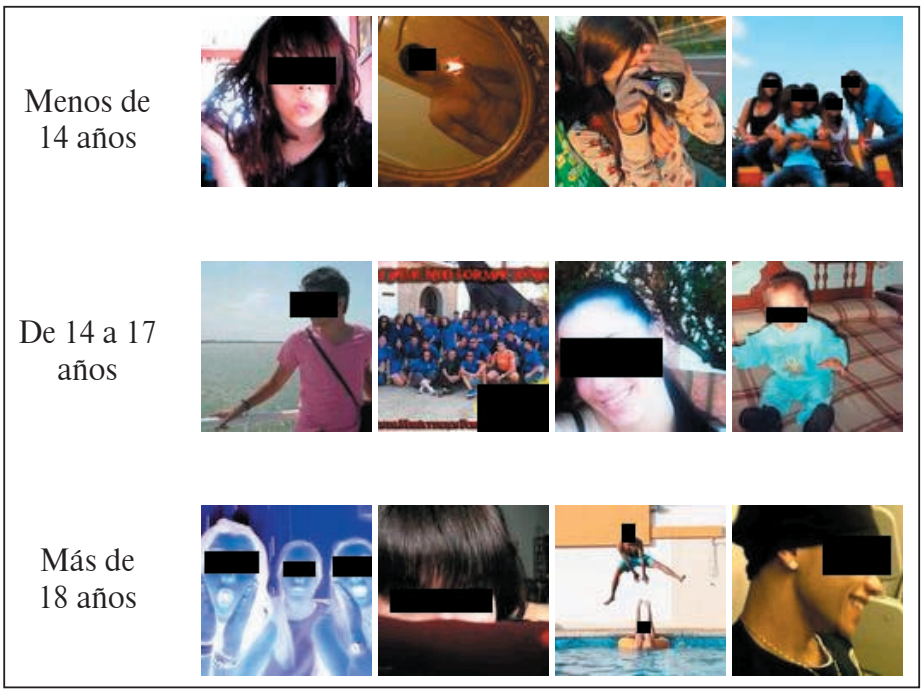

Figura 3. Ejemplos de fotografías tomadas de avatares en Tuenti.

4.3. Técnicas de detección de acoso sexual

El acoso sexual a menores o *grooming* es uno de los fenómenos más temidos por los progenitores, y también uno de los más frecuentes en las noticias periodísticas. A pesar de que su incidencia estadística es relativamente marginal en comparación con otros peligros en las redes sociales, el impacto social del fenómeno, así como el hecho de que médicamente se trate con frecuencia la pederastia[9] como una desviación sexual enfermiza, hace que el comportamiento de los pederastas en Internet haya sido estudiado en varios trabajos de investigación [20].

Aunque la interacción entre menores y pederastas a través de Internet, y a menudo a través de redes sociales, puede conllevar el intercambio de fotografías o grabaciones de vídeo, el medio principal de comunicación entre ambos es el chat o conversación escrita. El pederasta establece una relación continuada en el tiempo con el menor, a lo largo de la cual intenta averiguar información sobre el mismo, simulando inicialmente ser un menor de edad cercana, para progresivamente generar una relación que incluye intentos de desinhibición sexual, generación de dependencia y de culpa en el menor, chantaje, sexo virtual, y en último extremo, el establecimiento de contacto físico.

Dado que este tipo de comportamientos, así como el tipo de lenguaje empleado en ellos, es paradigmático, cabe distinguir dos tipos de enfoques para analizarlo:

■ Caracterizar los comportamientos de los pederastas a través de patrones léxicos y lingüísticos manualmente construidos para cada una de las conductas que se distinguen en el comportamiento de un pederasta.
■ Utilizar atributos léxicos y lingüísticos para representar los textos de pederastas y de usuarios que no lo son, y usar aprendizaje automático sobre esta representación con el fin de derivar automáticamente clasificadores que puedan distinguir unos de otros.

En ambos casos, para poder evaluar la efectividad de un enfoque, es necesario disponer de una colección de datos con conversaciones de pederastas y menores. Dado

el carácter de este tipo de conversaciones, es difícil conseguirlas ni siquiera para fines de investigación. Sin embargo, en el idioma inglés se dispone de un recurso inestimable, que es el sitio web *Perverted Justice*[10]. En este sitio se han hecho públicas gran número de conversaciones en chats entre pederastas convictos y voluntarios que previamente habían hecho de "ganchos", colaborando a su detención. Estas conversaciones, todas ellas en inglés, están disponibles para el público en general. En la **figura 4** se puede observar un fragmento bastante explícito de una de estas conversaciones. En ella, el acosador se identifica con el apodo "swgamaleyess", mientras que el voluntario que se hace pasar por menor del sexo femenino es "erin_lynnb".

En esta área, el trabajo presentado en [21] puede considerarse seminal. El autor utiliza una representación basada en trigramas de palabras para un problema de clasificación binario ("depredador vs. víctima"), y el algoritmo de aprendizaje kNN (k-*Nearest Neighbors*, los k vecinos más cercanos), para alcanzar un valor del F_1 de 94,3%.

En relación con los enfoques anteriores, el proyecto Chatcoder[11] del Ursinus College, liderado por la doctora April Kontostathis, es quizá la principal referencia. En el marco de este proyecto se han realizado diversas investigaciones, que inicialmente han utilizado modelos manualmente construidos del comportamiento de los pederastas [22] [23], y posteriormente modelos basados en aprendizaje (como en [24]). En este último trabajo, se ha simplificado el modelo de comportamiento usado en trabajos anteriores, y se ha etiquetado manualmente cada línea de una serie de conversaciones extraídas de *Perverted Justice* de acuerdo a una de las clases siguientes:

- Intercambio de información personal.
- Aproximación.
- Acoso.
- Ninguna de las anteriores.

Los autores del trabajo han desarrollado una representación basada en distintos atributos lingüísticos, como el uso de pronombres nominales, determinados nombres (e.g. "*age*", "*pic*", "*boyfriend*", "*penis*") y verbos característicos (e.g. "*teach*", "*suck*", "*kiss*"), etc. Sus pruebas con diversos algoritmos de aprendizaje alcanzan los mejores resultados con el algoritmo kNN, que obtiene un valor de exactitud (*accuracy*) de 82.4%. Hay que tener presente que no es un problema de distinguir toda la conversación de un pederasta de la de una víctima, sino clasificar cada frase dentro de una de las cuatro conductas anteriores, lo que es considerablemente más difícil.

Recientemente, en [25] se han utilizado cadenas léxicas para caracterizar uno de los posibles tipos de lenguaje de los pederastas, denominado por los autores "discurso fijado", y que consiste en abordar temas sexuales y desestimar los intentos del menor de cambiar de tema. Los experimentos demuestran que en los chats normales, se producen cambios rápidos de tema, mientras que en los chats de contenido sexual (ciber-sexo), el tema persiste durante toda la conversación. Sin embargo, en los chats de acosadores sexuales, el tema de las relaciones de este tipo se aborda paulatinamente.

Finalmente, cabe resaltar la existencia reciente de una competición científica relacionada con la detección de depredadores sexuales. En las series de competiciones PAN[12] (*Plagiarism Analysis, Authorship Identification, and Near-Duplicate Detection*), desarrolladas desde 2007 y ya en su séptima edición, se ha planteado una sub-tarea de la tarea de la identificación de autor, consistente en la detección de depredadores sexuales en chats, y las líneas de sus conversaciones que son identificativas de pederastia. A la fecha de la redacción de este artículo, los resultados de los competidores en esta tarea aún no se han publicado, pero cabe esperar que tanto las colecciones de datos como los enfoques probados se conviertan en la referencia fundamental en esta área.

```
swgamaleyess (07/15/06 9:56:30 PM): ur just lucky ur not
here
erin_lynnb (07/15/06 9:56:36 PM): y?
swgamaleyess (07/15/06 9:56:45 PM): i would be kissing u
erin_lynnb (07/15/06 9:56:55 PM): aww
swgamaleyess (07/15/06 9:57:05 PM): u ever had oral sex
erin_lynnb (07/15/06 9:57:10 PM): no
swgamaleyess (07/15/06 9:57:41 PM): so do u shave ur pussy
erin_lynnb (07/15/06 9:57:48 PM): didnt yet lol
swgamaleyess (07/15/06 9:58:16 PM): would u for me
erin_lynnb (07/15/06 9:58:27 PM): u want me to?
swgamaleyess (07/15/06 9:58:46 PM): before i taste u yes
```

Figura 4. Ejemplos de conversación extraída de *Perverted Justice*.

4.4. Técnicas de detección de acoso escolar

La detección del ciberacoso escolar (*cyberbullying*) es un problema a priori más complejo que los anteriores. En la detección de edad en redes sociales, existen múltiples fuentes de información que pueden utilizarse para construir modelos de comportamiento de menores y de adultos. En el caso del ciberacoso sexual, el fenómeno ha sido estudiado con detalle, se conocen las conductas típicas del comportamiento de un depredador sexual, y el fenómeno comienza siempre en el mundo *online*, donde se produce la toma de contacto.

Frente a estos peligros, el ciberacoso escolar raramente se produce en el mundo *online* exclusivamente, y menos aún comienza en este medio. Más bien al contrario, se trata de un fenómeno propio del entorno real del acosado, que tiene manifestaciones múltiples en el mundo *online*. Algunos de los comportamientos relacionados con este fenómeno que se dan en redes sociales son los siguientes:

- Insultos y vejaciones textuales.
- Difusión de fotos comprometedoras, o de fotos modificadas y vergonzosas de la víctima.
- Toma y difusión de vídeos vejatorios, como el "*happy slapping*"[13] y otros tipos de humillaciones públicas.
- Exclusión de grupos sociales, como el grupo de amigos de la clase escolar de la víctima en la red social.
- Suplantación de la víctima en distintos foros, exhibiendo conductas comprometedoras y auto-excluyentes.
- Denuncia de la víctima ante los propietarios del sistema *online* para que su cuenta sea cancelada.

Es obvio que el diseño de técnicas semiautomáticas para la detección de cada una de estas variedades es un problema complejo, y algunas de ellas quizá nunca puedan ser desarrolladas.
Actualmente, los investigadores han concentrado fundamentalmente sus esfuerzos en el análisis de interacciones textuales (chats, comentarios en muros, etc.) con el fin de caracterizar y detectar la primera de las variedades anteriormente citadas. En este sentido, existen varios trabajos relevantes:

- En [26] se ha construido una colección de datos usando preguntas y respuestas de la red social Formspring.me, etiquetadas manualmente con presencia de *cyberbullying* o no, y se ha construido un clasificador usando aprendizaje automático (con árboles de decisión usando C4.5) sobre una representación de los comentarios utilizando las frecuencias y proporciones de aparición de una serie de insultos y palabras soeces. El sistema es capaz de detectar un 67,4% de los comentarios de acoso, y puede alcanzar un 81,7% de efectividad total.

■ En [27] se plantea el problema de la detección de ciberacoso en comentarios de Youtube. Se ha construido un clasificador de dos niveles, en el que el primer nivel determina si el comentario está relacionado con temas potencialmente sensibles (sexualidad, raza y cultura, inteligencia, atributos físicos), y el segundo nivel determina cual es el tema en particular. Los autores han conseguido alcanzar una eficacia del 66.7% usando un clasificador basado en SVM.
■ En [28] se presenta un modelo teórico de representación de emociones en textos, orientado a representar en lenguaje XML tanto las emociones como los comportamientos de carácter ofensivo o agresivo en interacciones textuales. No se construye ninguna colección ni se evalúa ningún enfoque técnico.

Como se puede observar, incluso en el caso del análisis textual, los trabajos existentes son escasos, e incluso se puede argumentar que los ejemplos usados en las colecciones de entrenamiento y evaluación no son exactamente de ciberacoso escolar, sino del concepto más genérico del comportamiento ofensivo entre personas (posiblemente adultos) a través de Internet. Por tanto, se puede afirmar que aún queda mucho camino por andar en este ámbito.

4.5. Técnicas de detección de pornografía infantil

Como ya hemos comentado previamente, existen evidencias pero no estadísticas sobre la existencia de anillos o grupos de pedófilos que intercambian imágenes sexuales más o menos explícitas en redes sociales. Cabe reseñar que los pedófilos siguen utilizando otros entornos para el intercambio de material ilegal, como las redes P2P, determinados foros de los grupos de noticias (*newsgroups*), determinados chats, etc. En particular, existen evidencias recientes de que los pedófilos están usando múltiples métodos de anonimización para acceder a la "Web oculta" ("*hidden Internet*")[14], es decir, sitios web que sólo pueden ser accedidos usando software de anonimización como la red Tor[15].

En cualquier caso, a fecha de hoy existen dos enfoques para la detección de pedófilos en Internet:

■ La utilización de palabras clave características (e.g. "*boylover*", "*lolita*", "*teen sex*", etc.) para la realización de búsquedas en el entorno objetivo (la red social, P2P, etc.). Esto supone la búsqueda activa por medio de humanos que deben examinar luego los resultados y decidir si el contenido compartido es ilegal o no, o bien si viola los Términos del Servicio.
■ La utilización de bases de datos de imágenes y vídeos de pornografía infantil y su uso como material de comparación haciendo búsquedas por "*hash*" o identificador unívoco basado en el contenido de los archivos. Esta técnica es habitual en redes P2P, donde cada archivo está identificado por un "*hash*" similar a una firma MD5, de modo que un programa puede buscar la presencia de los *hashes* de los archivos de imágenes pedófilas conocidas en las carpetas compartidas de los usuarios de P2P.

Esta segunda técnica es la que utilizan diversas policías y cuerpos de seguridad, como la Interpol, la Policía Nacional, o la Guardia Civil. Conviene recordar que en la mayoría de las legislaciones, la simple tenencia de este material es constitutivo de delito, por lo que dicha técnica está reservada sólo a aquellos que pueden disponer de dicho material de manera legal. Por ejemplo, la Guardia Civil utiliza buscadores de material pedófilo en las redes P2P como Híspalis o Nautilus[16].

Una de las principales deficiencias de este último tipo de tecnologías es que, como en cualquier tipo de "*hash*", la modificación de un simple bit del archivo lleva a la obtención de un "*hash*" o firma sustancialmente distinta. Por tanto, un pedófilo puede compartir material marcado efectuando triviales modificaciones en el mismo, como cambiar de color un simple píxel.

Por el contrario, las técnicas de búsqueda basadas en palabras clave características, aunque menos efectivas y más laboriosas, son resistentes a modificaciones.

En este ámbito, el proyecto "*Measurement and Analysis of P2P Activity Against Paedophile Content*"[17], desarrollado por varios organismos de investigación y universidades europeas bajo los auspicios del *Programa Safer Internet Plus* de la Comisión Europea y de otras entidades, tiene una especial relevancia. En este proyecto se ha desarrollado de manera semiautomática una base de datos de palabras clave características de las búsquedas pedófilas, usando las búsquedas observadas en la red EDonkey y clasificándolas como búsquedas candidatas a partir de un conjunto semilla de palabras clave pedófilas conocidas previamente. Estas consultas han sido validadas posteriormente por medio de expertos en el problema. El propio protocolo de construcción y validación constituye una de las aportaciones principales, puesto que la idea es clasificar las búsquedas realizadas en los sistemas P2P como pedófilas o no en el tiempo [29]. La lista de consultas está disponible en el servidor del proyecto, pero requiere permiso especial.

5. Conclusiones

En este artículo hemos revisado los principales peligros que acechan a los menores en las redes sociales. Como es lógico, por su propio carácter las redes sociales pueden tener un efecto catalizador, pero no promueven la existencia de peligros más allá de los previamente existentes. En todo caso, la presencia de menores en estas redes puede haber afectado más bien a la incidencia de determinados peligros. Por ejemplo, tradicionalmente se ha considerado la exposición a contenidos inapropiados como el peligro más frecuente al que se enfrentan los menores en Internet. Sin embargo, en el ámbito de las redes sociales, la incidencia de este peligro es notablemente inferior a otros ámbitos, porque los menores no suelen ni crear ni compartir este tipo de materiales, salvo en algunos casos especiales (e.g. blogs pro-anorexia y pro-bulimia).

Como resultado de la presión política y social, y a la vista de que las medidas legales resultan inefectivas en muchos casos, los proveedores y operadores de redes sociales han tomado una serie de medidas de auto-regulación que se pueden considerar, con carácter general, como bastante reactivas. En otras palabras, el personal de revisión o moderación de la red social interviene bajo demanda cuando los usuarios denuncian o notifican la existencia de un material o de una situación dañina o ilegal. En los propios principios adoptados por los operadores se reconoce la necesidad de desarrollar métodos técnicos efectivos para la detección de los peligros que afectan a los menores en las redes sociales.

Hemos revisado con bastante detalle el estado de las técnicas que permiten detectar bien la edad de los usuarios de una red social, o bien determinados peligros como el acoso sexual, el acoso escolar o la pornografía infantil. Como balance general de las técnicas revisadas, se puede afirmar que existen muchos resultados prometedores en la literatura científica reciente, que pueden dar lugar en un relativo corto plazo a sistemas efectivos que ayuden a los operadores de redes sociales a adoptar una actitud más proactiva. En particular, las técnicas que se han revelado más eficaces son las relacionadas con la clasificación de texto basada en aprendizaje, pero frente a representaciones más tradicionales basadas en palabras clave, en estos casos resultan especialmente efectivas las representaciones de textos que utilizan secuencias de palabras (e.g. trigramas) y toda una serie de atributos léxicos y lingüísticos. Este tipo de técnicas podrían usarse de manera semiautomática, como filtro para priorizar las revisiones realizadas por expertos humanos, o para alertar a los mismos de los perfiles a revisar.

Agradecimientos

Esta investigación ha sido financiada parcialmente por Optenet S.A. y por el Ministerio de Economía y Competitividad y el Centro para el Desarrollo Tecnológico Industrial (CDTI), en el marco del proyecto de investigación industrial "*WENDY: WEb-access coNfidence for chilDren and Young*" (TSI-020100-2010-452).

Referencias

[1] European Commission. *Public Consultation on Online Social Networking, Summary Report*, 2008. <http://ec.europa.eu/information_society/activities/sip/public_consultation/index_en.htm>.

[2] INTECO. *Estudio sobre hábitos seguros en el uso de las TIC por niños y adolescentes y e-confianza de sus padres*. Observatorio de la Seguridad de la Información (INTECO), marzo 2009. <http://www.inteco.es/file/O4-7X0FfwOb7HFjd-HHpx7Q>.

[3] A. Lenhart, M. Madden. Teens, Privacy & Online Social Networks. *Pew Internet & American Life Project, 2007*. <http://www.pewinternet.org/Reports/2007/Teens-Privacy-and-Online-Social-Networks.aspx>.

[4] J. Palfrey. *Enhancing Child Safety & Online Technologies*. Berkman Center for Internet & Society at Harvard University, 2008. <http://cyber.law.harvard.edu/pubrelease/isttf/>.

[5] A. Lenhart, K. Purcell, A. Smith, K. Zickuhr. Social Media and Young Adults. *Pew Internet & American Life Project, 2010*. <http://pewinternet.org/Reports/2010/Social-Media-and-Young-Adults/Summary-of-Findings.aspx>.

[6] S. Livingstone et al. *EU Kids Online Final Report*, 2011. The London School of Economics and Political Science. <http://www.eukidsonline.net>.

[7] M. Garmendia et al. *Riesgos y seguridad en internet: Los menores españoles en el contexto europeo*, 2011. The London School of Economics and Political Science. <http://www.ehu.es/eukidsonline>.

[8] J.M. Gómez Hidalgo, E. Puertas Sanz, F. Carrero, M. de Buenaga Rodríguez. Web Content Filtering. En (ed.) Marvin V. Zelkowitz, *Advances in Computers 76*, pp. 257-306, Elsevier Academic Press, 2009.

[9] J.M. Gómez Hidalgo, G. Cánovas Gaillemín, J.M. Martín Abreu. Medidas técnicas de protección del menor en Internet. *Novática nº 209*, enero-febrero 2011, pp. 42-48.

[10] Y. Azuma, M. Nishimoto, N. Miyamoto, T.X. Fujisawa, N. Nagata, A. Kosaka. A comparative assessment of one's own age from facial images of others: Two case studies for the americans and the japanese. *IEEE International Conference on Systems, Man and Cybernetics (SMC 2009)*.

[11] Y. Fu, T.S. Huang. Human Age Estimation with Regression on Discriminative Aging Manifold. *IEEE Transactions on Multimedia 2008*, pp. 578-584.

[12] X. Geng, Z.-H. Zhou, Y. Zhang, G. Li, H. Dai. Learning from facial aging patterns for automatic age estimation. *Proceedings of the ACM Conference on Multimedia, 2006*, pp. 307–316.

[13] A. Lanitis, C. Draganova, C. Christodoulou. &RPSDULQJ GLIIHUHQW FODVVL HUV IRU DXWRPDWLF DJH estimation. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, v. 34, n. 1*, 2004, pp. 621–628.

[14] N. Ramanathan, R. Chellappa.)DFH YHUL FDWLRQ across age progression. *IEEE Transactions on Image Processing, v. 15, n. 11*, 2006, pp. 3349–3361.

[15] A. de Santos Sierra, C. Sánchez Ávila, M. Carmonet Bravo, J. Guerra Casanova, D. de Santos Sierra. Control de edad en redes sociales mediante biometría facial. *XII Reunión Española sobre Criptología y Seguridad de la Información (RECSI 2012)*, Donostia-San Sebastián.

[16] L.A.F. Mendoza, E. Cataldo, M. Vellasco, M.A. Silva, A.D.O. Cañandon, J.M. de Seixas. Classification of voice aging using ANN and glottal signal parameters. *Proceedings of IEEE ANDESCON 2010*, pp. 1–5.

[17] J. Lu, Y-P. Tan. Gait-based human age estimation. *IEEE Transactions on Information Forensics and Security, v. 5, n. 4*, 2010, pp. 761–770.

[18] C. Peersman, W. Daelemans L. Van Vaerenbergh. Age and Gender Prediction on Netlog Data. *Proceedings of the Twenty first Meeting of Computational Linguistics in the Netherlands (CLIN21)*, 2011, Ghent, Bélgica.

[19] J. Tam, C. Martell. Age Detection in Chat. *Proceedings of the Third IEEE International Conference on Semantic Computing*, 2009, Berkeley, EE.UU.

[20] L. Olson, J. Daggs, B. Ellevold, T. Rogers. Entrapping the innocent: Toward a theory of child sexual predators' luring communication. *Communication Theory 17, n. 3*, 2007, pp. 231-251.

[21] N. Pendar. Toward Spotting the Pedophile Telling victim from predator in text chats. *International Conference on Semantic Computing (ICSC '07)*. IEEE Computer Society, Washington, DC, EE.UU., pp. 235-241.

[22] A. Kontostathis, L. Edwards, A. Leatherman. ChatCoder: Toward the Tracking and Categorization of Internet Predators. *Proceedings of the Text Mining Workshop 2009*, Sparks, Nevada, EE.UU.

[23] A. Kontostathis, L. Edwards, J. Bayzick, I. McGhee, A. Leatherman K. Moore. Comparison of Rule-based to Human Analysis of Chat Logs. *First International Workshop on Mining Social Media (MSM09)*, Sevilla, España.

[24] I. McGhee, J. Bayzick, A. Kontostathis, L. Edwards, A. McBride, E. Jakubowski. Learning to Identify Internet Sexual Predation. *International Journal on Electronic Commerce, v. 15, n. 3*, 2011.

[25] D. Bogdanova, P. Rosso, T. Solorio. Modelling Fixated Discourse in Chats with Cyberpedophiles. *Proceedings of the Workshop on Computational Approaches to Deception Detection, Association for Computational Linguistics, April 2012*, Avignon, Francia, pp. 86-90.

[26] K. Reynolds, A. Kontostathis, L. Edwards. Using Machine Learning to Detect Cyberbullying. *Proceedings of the 10th International Conference on Machine Learning and Applications Workshops* (ICMLA). December 2011. Honolulu, Hawaii, EE.UU.

[27] K. Dinakar, R. Reichart, H. Lieberman. Modeling the Detection of Textual Cyberbullying. *Fifth International AAAI Conference on Weblogs and Social Media* (SWM'2011), Barcelona, España.

[28] M. Munezero, T. Kakkonen, C. Montero. Towards Automatic Detection of Antisocial Behavior from Texts. Fifth International Joint Conference on Natural Language Processing, 2011, Chiang Mai, Tailandia.

[29] M. Latapy, C. Magnien, R. Fournier. Quantifying paedophile queries in a large P2P system. *Proceedings of INFOCOM 2011*: pp. 401-405.

Notas

[1] Entrevista a Guillermo Cánovas, Presidente de Protégeles: <http://www.fundacionctic.org/actualidad-y-recursos/sala-de-prensa/la-mayoria-de-amigos-de-las-redes-sociales-no-les-podriamos-pedir-un-favor>.

[2] Se pueden revisar diversas medidas legislativas a nivel europeo en: <http://ec.europa.eu/information_society/activities/sip/policy/legislation/index_en.htm>.

[3] Tanto la declaración de principios, como los signatarios de la misma, están disponibles en: <http://ec.europa.eu/information_society/activities/sip/self_reg/social_netwk/index_en.htm>.

[4] *Digital Agenda: only two social networking sites protect privacy of minors' profiles by default*. <http://europa.eu/rapid/pressReleasesAction.do?reference=IP/11/762>.

[5] *Digital Agenda: social networks can do much more to protect minors' privacy.* <http://europa.eu/rapid/pressReleasesAction.do?reference=IP/11/1124>.

[6] <http://europa.eu/rapid/pressReleasesAction.do?reference=IP/12/445>.

[7] Este proyecto de I+D, en el que colaboran la empresa Optenet y el Centro de Domótica Integral (CeDInt) de la UPM, se encuentra actualmente en desarrollo y la única publicación disponible en este momento sobre el mismo es [15], centrada en análisis facial.

[8] <http://www.fgnet.rsunit.com/>.

[9] Cabe distinguir entre los pedófilos, o sexualmente inclinados hacia a los menores, y los pederastas, que activamente persiguen y con cierta frecuencia logran establecer contactos sexuales de esta naturaleza.

[10] <http://perverted-justice.com/>.

[11] <http://www.chatcoder.com/>.

[12] <http://pan.webis.de/>.

[13] Un conjunto de menores golpean a la víctima con palmadas en la parte anterior del cuello ("collejas"), mientras lo graban con un teléfono móvil.

[14] The Telegraph. "*Police target 'anonymous' paedophiles on 'hidden internet'*". <http://www.telegraph.co.uk/technology/news/9345144/Police-target-anonymous-paedophiles-on-hidden-internet.html>.

[15] <https://www.torproject.org/>.

[16] <http://www.guardiacivil.es/es/prensa/noticias/historico2/2805.html>.

[17] <http://antipaedo.lip6.fr/>.