

PID 7044 – “Gestión de Datos Biométricos en Bases de Datos Objeto-Relacionales”

Director: Carlos Eduardo Alvez

Codirector: Etchart, Graciela Raquel

Integrantes Docentes: Aguirre, Juan José; Benedetto, Marcelo Gabriel; Miranda, Ernesto Martín; Ruiz, Silvia Estela

Resumen

La aplicación de sistemas biométricos, particularmente en la administración pública, ha tomado importancia en los últimos años, tanto para controles de accesos físicos o virtuales, documentación, etc. El uso de esta tecnología no alcanza si no se cuenta con bases de datos sobre las cuales se pueda contrastar la información que brindan las aplicaciones basadas en biometría. Para esto, es importante que los organismos estatales que utilicen grandes volúmenes de datos biométricos, puedan consultar la información de manera eficiente e intercambiar la misma con otros organismos asociados. De allí la necesidad de contar con estándares que permitan la interoperabilidad entre sistemas y así facilitar la búsqueda de datos. Para esto, en este proyecto que da continuidad al Proyecto PID 07/G035 “*Identificación de personas mediante Sistemas Biométricos. Estudio de factibilidad y su implementación en organismos estatales*”, se propone la gestión de datos biométricos en Bases de Datos Objeto-Relacionales, permitiendo la representación tanto de los datos crudos (imágenes/audios) como los metadatos requeridos por estándares utilizados en la comparación, e indexación para un acceso eficiente a los mismos. También, se crearán modelos de representación de la información necesaria para crear registros de transacciones ANSI-NIST-ITL-1 2011 [1], y de esta manera tener la posibilidad de intercambiar información entre organismos asociados, mediante procesos de exportación e importación.

Palabras clave: Bases de datos OR, biometría, organismos públicos.

Introducción

La biometría es una tecnología que hoy en día se encuentra en pleno desarrollo, tanto en el ámbito de la vida cotidiana como en la investigación. Esta rápida evolución se debe a la creciente preocupación por la seguridad y por la vinculación que tiene esta tecnología para garantizar la misma. En cuanto a la vida cotidiana, desde una mirada macro del sector, la Argentina es un país con mucho potencial en la región en implementar y desarrollar soluciones de biometría tanto para el sector público como privado [6] [7]. La mayor importancia en el uso de esta tecnología radica en su aplicación como mecanismo de control: control de acceso físico, acceso virtual (redes, sistemas, etc.), controles de horarios, entre otros.

Los organismos públicos que emplean esta tecnología necesitan consultar los datos biométricos de manera eficiente e intercambiar la misma con otros organismos asociados. De allí la necesidad de contar con modelos de datos y con estándares que permitan la interoperabilidad entre sistemas y así facilitar la búsqueda de datos.

Uno de los estándares que contempla este tipo de aspectos (entre otros), es el ANSI/NIST ITL 1-2011 [1], norma biométrica publicada en noviembre de 2011, que define cómo trabajar para garantizar la interoperabilidad de datos biométricos entre los distintos sistemas. Este estándar, define el contenido, el formato y las unidades de medida para el intercambio de huellas dactilares, palmares, plantares, faciales, el iris, el ácido desoxirribonucleico (ADN), y otras muestras biométricas y datos forenses que pueden ser utilizados en el proceso de identificación o verificación de una persona. Es el estándar más utilizado por entes estatales, y sobre el mismo se viene trabajando en el grupo de

investigación del proyecto PID 07/G035, específicamente en el rasgo de iris (registro 17 del estándar). Existe un borrador para voz (registro 11) [8], lo cual es importante porque es uno de los rasgos sobre el que se trabajará en el presente proyecto.

Para que las muestras biométricas capturadas por un organismo determinado, puedan ser intercambiadas con otros organismos, el sistema que gestiona los datos biométricos debe soportar registros de transacciones conforme a esta norma, es decir, debe ser capaz de generar y/o utilizar las transacciones que sean morfológicas, sintácticas y semánticamente conformes a los requisitos del estándar. Para esto, el sistema debe contar con los módulos de software que se encarguen tanto de la recepción de estos registros, como también, de la elaboración de los mismos para ser enviados a organismos asociados [4].

Los registros antes mencionados, se utilizan únicamente para el intercambio de información. Sin embargo, se debe también considerar la generación y el almacenamiento de metadatos necesarios para el reconocimiento, o sea para los procesos de identificación y/o verificación de personas. Como ejemplo de estos metadatos se pueden mencionar: las minucias de una huella dactilar, o el código del iris (*iriscode*) [9][10], en un sistema de reconocimiento del iris, rasgos acústicos para el reconocimiento de voz, entre otros. O sea, que estos sistemas tienen que tratar, por un lado con: imágenes 2D (iris, huellas dactilares, etc.), imágenes 3D (reconocimiento facial, geometría de la mano, etc.), sonidos (reconocimiento de voz), videos (reconocimiento de gestos o movimientos corporales), y por otro lado, la gestión de metadatos extraídos de estos datos para ser utilizados en el reconocimiento (minucias de huellas, código de iris, etc.) que son utilizados para la comparación. Estos últimos son estructuras complejas, y pueden ocasionar algunos inconvenientes si se trabaja con el modelo de datos relacional, por las limitaciones impuestas por el mismo. Por esto, aquí se ha optado por utilizar la tecnología Objeto-Relacional (OR), como se detallará en la próxima sección [5] [11].

Líneas de Investigación y Desarrollo

En la actualidad, en el área específica de la biometría, se encuentran abiertas varias líneas de investigación con distintos niveles de desarrollo. Sin embargo, como se trató en la sección anterior, especialmente en organismos gubernamentales, es muy importante considerar el nivel de estandarización alcanzado, sobre todo, en lo referente al formato de intercambio de datos. Por lo tanto, se debe tener particularmente en cuenta esto, para el desarrollo de sistemas biométricos en organismos públicos [2] [3].

Otra línea de investigación vigente está orientada a sistemas que combinan diferentes rasgos biométricos o multimodales. Estos sistemas son más precisos y seguros dado que superan algunas limitaciones de los sistemas que utilizan un único rasgo biométrico. Las limitaciones de un rasgo pueden ser: no universalidad, ruidos en los datos, suplantación de identidad, entre otros [12].

Sin embargo, una de las líneas más importantes, y la que nos interesa aquí, está relacionada con la construcción de modelos para la representación de datos relacionados con los datos biométricos, de manera que los mismos puedan consultarse, recuperarse y compararse de manera simple y eficiente (el método de comparación “*matching*” se utiliza en los procesos de identificación/verificación). Esto además respetando las especificaciones de los estándares adoptados.

Estos modelos, tienen estructuras de datos complejas y pueden ocasionar algunos inconvenientes si se trabaja con el modelo de datos relacional:

1. Las estructuras de los metadatos para la búsqueda y comparación de datos biométricos (*iriscode*, minucias, etc.), por lo general son estructuras no atómicas, por lo que en las bases de datos relacionales (para respetar la primera forma normal) se deben separar en varias tablas. Esto hace que sea menos eficiente el tratamiento. Por ejemplo, en la comparación de dos conjuntos de metadatos, necesita realizar operaciones de *join*.

2. El procesamiento de los datos se debe hacer desde lenguajes particulares (C++, java, etc.), lo que presenta el problema de la compatibilidad (y dependencia) con los lenguajes de programación, ya que se deben transferir los datos generados por métodos implementados en distintos lenguajes a la base de datos para ser almacenados. Este tema no es trivial, y se debe tratar para cada lenguaje en particular.
3. Relacionado con lo anterior, para que un lenguaje pueda tratar con los datos, como por ejemplo con el método de comparación entre dos rasgos biométricos, se deben crear cursores que comúnmente deben ser transferidos a través de una red para que retornen un único valor. Esto puede provocar problemas de tráfico de red innecesarios, dado que estos procesos retornan un único valor lógico, en el caso de la comparación, utilizado en el proceso de verificación, retorna si verifica o no verifica.

Frente a estos inconvenientes de las bases de datos relacionales, aquí se ha optado por utilizar el modelo Objeto-Relacional (OR) [12]. Este modelo permite dar soluciones a las limitaciones de las bases de datos relacionales ya que brindan las siguientes facilidades:

1. **Definir tipos de datos:** estos tipos pueden contener estructuras complejas como colecciones, objetos grandes, etc. sin la limitación de la primera forma normal.
2. **Definir e implementar el comportamiento de los datos:** se pueden crear los métodos que gestionen los datos de los tipos definidos, y así facilitar el acceso seguro desde las aplicaciones que los utilicen.
3. **Definir e implementar métodos de acceso de dominio:** esto permite mejorar los tiempos de acceso en la recuperación de datos de un dominio específico, como en el caso de los datos biométricos.

Las facilidades mencionadas, permiten crear infraestructuras que extienden los servicios de del Sistema de Gestión de Base de Datos [13]. Estas le permiten a este sistema tratar con datos de un dominio específico, en este caso, con datos biométricos.

En lo que concierne al proyecto, esta infraestructura debe contemplar:

- **La creación de tipos específicos para los datos biométricos:** esto incluye tipos para los datos crudos (imágenes, audios, videos), tipos para la codificación de los mismos (*iriscode*, minucias, etc.), y todos los metadatos necesarios para generar los registros de transacciones ANSI/NIST-ITL1-2011.
- **Definir e implementar los métodos que gestionen los tipos antes definidos:** aquí se incluyen, entre otros, la generación de las codificaciones específicas, métodos de comparación (*matching*), métodos de generación e importación de registros ANSI/NIST-ITL1-2011.
- **Crear índices de dominio que permitan mejorar los tiempos de respuestas en los procesos de identificación:** en la identificación biométrica, se necesita acceder a varios registros para poder identificar a una persona. Mejorar la eficiencia en estos accesos, no es simple de realizar con índices tradicionales como árboles b, hash, etc. Esto se debe a que los códigos (plantillas) generadas para la comparación, son datos multidimensionales y necesitan métodos de acceso específicos para el dominio.

Objetivos

El objetivo general de este trabajo es desarrollar una infraestructura genérica en BDOR para la identificación de personas mediante el reconocimiento de iris y voz que permita la interoperabilidad entre organismos.

Como objetivos específicos se tienen:

- Realizar los ajustes para adecuar a normas internacionales el software para reconocimiento mediante iris desarrollado en el proyecto anterior.
- Implementar un software para el reconocimiento de voz acorde a normas internacionales.
- Crear modelos en BDOR, que posibiliten almacenar tanto las plantillas como los metadatos necesarios para adecuar los formatos de intercambio de datos al estándar (ANSI/NIST-ITL 1-2011), de manera que permita compartir información con diferentes organismos.
- Proponer métodos de acceso en BDOR para la mejora de la eficiencia en los procesos de identificación de iris y/o voz.
- Crear prototipos utilizando dispositivos estándares y de bajo costo para la captura de la imagen del iris y voz.

Referencias

1. Wing B. ANSI/NIST-ITL 1-2011. Information Technology: American National Standard for Information Systems Data Format for the Interchange of Fingerprint, Facial & Other Biometric Information. November, 2011.
2. Carlos E. Alvez, Marcelo G. Benedetto. “Los Sistemas Biométricos y su Factibilidad de Aplicación en los Organismos Estatales”. XII Workshop de Investigadores en Ciencias de la Computación (WICC 2010), El Calafate, Santa Cruz Argentina, 5 y 6 de Mayo de 2010, Universidad Nacional de la Patagonia San Juan Bosco. Páginas 247-251.
3. Graciela Etchart, Lucas Luna, Carlos Leal, Marcelo Benedetto, Carlos Alvez. Sistemas de reconocimiento biométricos, importancia del uso de estándares en entes estatales. CGIV - XIII Workshop de Investigadores en Ciencias de la Computación (WICC 2011), 5 y 6 de Mayo de 2011. Universidad Nacional de Rosario. Rosario – Argentina. Páginas 339-343.
4. Graciela Etchart, Lucas Luna, Rafael Leal, Marcelo Benedetto, Carlos Alvez. “Sistema adecuado a estándares de reconocimiento de personas mediante el iris”. CGIV - XIV Workshop de Investigadores en Ciencias de la Computación (WICC 2012), 25 y 26 de Abril de 2012. Universidad Nacional de Misiones. Posadas – Argentina. Páginas 321-325.
5. Alvez Carlos E. Modelos para la recuperación de imágenes por similitud en Bases de Datos Objeto-Relacionales. Tesis Doctoral. Santa Fe, Argentina, Abril de 2012. ISBN 978-987-33-2249-5.
6. Casal Gabriel, Rovolta Mercedes. Biometrías. Herramientas para la Identidad y la Seguridad Pública. Jefatura de Gabinete de Ministros. Presidencia de la Nación. Noviembre de 2010.
7. Julio Fuoco, Tendencias Biométricas, desafíos y oportunidades. En Biometrías 2. Jefatura de Gabinete de Ministros. Presidencia de la Nación. Octubre de 2011.
8. Development of Draft Type-11 Voice Signal Record. Investigatory Voice Biometrics Committee Report, 9 March, 2012. In http://biometrics.nist.gov/cs_links/ibpc2012/presentations/Day4/IVBCReport_v1.8.pdf.
9. Daugman, J.: High condence visual recognition of persons by a test of statistical independence. IEEE Transactions on Pattern Analysis and Machine Intelligence 15(11), 1148-1161 (November 1993).
10. J. Daugman and C. Downing, “Effect of severe image compression on iris recognition performance,” IEEE Transactions on Information Forensics and Security, vol. 3, no. 1, p. 52–61, 2008.

11. Melton Jim, “(ISO-ANSI Working Draft) Foundation (SQL/Foundation)”, ISO/IEC 9075-2:2003 (E), United States of America (ANSI), 2003.
12. Alvez C., Benedetto M., Berón G., Etchart G., Luna L. y Leal C. Desarrollo de un sistema multi-biométrico mediante reconocimiento de iris y voz, adecuado a estándares, para su aplicación en organismos públicos. SIE 2011 – Simposio de Informática en el Estado. Córdoba, 31 de Agosto, 01 y 02 de Septiembre de 2011. 40° JAIIO. Páginas: 206 - 220.
13. Carlos E. Alvez, Aldo R. Vecchietti. Combining Semantic and Content Based Image Retrieval in ORDBMS. KES’2010, Lecture Notes in Computer Science, 2010, Volume 6277/2010, 44-53. Editors Rossitza Setchi. Ivan Jordanov, Robert J. Howlett, Lakhmi C. Jain. Springer-Verlag Berlin Heidelberg (2010).