

Desarrollo de un Índice Maestro de Pacientes Utilizando Estándares y Software Open Source

Carlos Alejandro Martínez¹, Ricardo Cimarelli³, Tamara Fazio², Gerardo Fuentes⁴

¹ Universidad Tecnológica Nacional-Regional Mendoza. Departamento de Sistemas de Información. carlos.martinez@frm.utn.edu.ar

² Hospital Regional Dr. Antonio J. Scaravelli. Comité de Docencia e Investigación. tgfazio@mendoza.gov.ar

³ Hospital Dr. Victorino Tagarelli. Informática. ricardo.cimarelli@gmail.com

⁴ Dirección de Informática del Ministerio de Salud de la Provincia de Mendoza. Departamento de Proyectos. gafuentes@mendoza.gov.ar

Resumen. Es fundamental, que todo efector relacionado con la salud que tenga sistemas informáticos implemente procesos que permitan identificar en forma única a una persona evitando los problemas relacionados con la duplicación de pacientes. Por este motivo, en este trabajo se presenta una propuesta de desarrollo de un Índice Maestro de Pacientes (IMP) utilizando estándares y centrado en la implementación del módulo de búsquedas y auditoría. El objetivo fundamental es definir una arquitectura utilizando software open source que permita realizar los procesos necesarios para lograr el objetivo de evitar y corregir la duplicación de los datos de los pacientes aumentando la calidad de estos.

Palabras Claves: Índice Maestro de Pacientes, Búsqueda de Candidatos.

1 Introducción

En el marco de la implementación de la historia clínica compartida es indispensable resolver el problema de la duplicación de los registros de identidad de los pacientes porque esto puede tener consecuencias negativas en la salud de los pacientes. La identificación errónea de personas puede producir la pérdida de información o mala praxis. Como se explica en [1] las consecuencias pueden ser por ejemplo: la no detección de alergias medicamentosas que puede llevar a un error médico.

Las duplicaciones se producen comúnmente por errores humanos en la entrada manual de datos que no son detectados por los sistemas de información y por la posible multiplicidad de números de documentos como la cédula o el documento identidad. Comúnmente los sistemas utilizan como identificador único de personas a números como el documento nacional de identidad. Está probado que estos identificadores no son suficientes para identificar unívocamente una persona [2].

Es común el caso donde un conjunto de organismos de salud dependientes de una entidad central utilizan cada uno su propio identificador de paciente sin existir control sobre la duplicación de los datos de identificación relacionados con este. Estos regis-

tros con los datos básicos de las personas luego abren la posibilidad de que existan historias clínicas duplicadas. Tiene que ser una prioridad de cualquier efector de salud mantener la integridad de la datos de filiación de los pacientes para tener una base fiable sobre la que funcione la historia clínica electrónica.

Entre organismos de salud se da el caso que cada entidad gestiona su propios datos de filiación, sin tener control de la duplicidad de dichos datos en los sistemas de otros organismos de salud, ya que no existe comunicación entre ellos. Esta falta de comunicación provoca inconsistencias de información entre las aplicaciones. Una misma persona puede estar dada de alta en diferentes organismos, y sus datos de filiación pueden variar de uno a otro.

El problema se resume en las siguientes líneas:

- Se dispone de un conjunto de efectores de salud con sistemas totalmente aislados. No existe comunicación entre ellos.
- En cada sistema se registran datos de pacientes utilizando como identificador un número o una cadena de caracteres alfanumérica.
- Inconsistencias de información entre las aplicaciones de distintos efectores.
- No existen controles sobre la calidad de los datos de filiación de los pacientes. Existe la posibilidad de duplicación de estos datos.

Teniendo en cuenta la problemática expuesta surge como una prioridad el contar con un sistema de identificación único de pacientes. Este debe permitir identificar a una persona en forma única en todo el sistema de salud y evitar la duplicación de registros que identifican a las personas; mediante la implementación de una serie de procesos. Esto debe funcionar en todo el sistema de salud y se deben proveer los mecanismos de auditoría necesarios para su funcionamiento.

El presente trabajo tiene como objetivo presentar la propuesta de desarrollo open source de un Índice Maestro de Pacientes, implementada en el Hospital Antonio J. Scaravelli[4] de la provincia de Mendoza, República Argentina. El desarrollo implementa criterios estrictos para evitar el ingreso de posibles duplicaciones y mecanismos de Auditoría del padrón de pacientes para el mantenimiento de la calidad de los datos. En los puntos siguientes, se describen los elementos utilizados para el desarrollo y los aspectos centrales para la implementación de todo el proceso. Este trabajo es tomado como base para el actual desarrollo del índice maestro de pacientes de la provincia de Mendoza.

2 Elementos del Trabajo y Metodología

2.1 Metodología y Relación con Otras Propuestas

La metodología de trabajo utilizada se basa en algunos de los elementos planteados en los siguientes trabajos: [5] y [6]. En base al objetivo planteado para el desarrollo de un Índice Maestro de Pacientes, se estudian todos los temas teóricos necesarios como los detallados en [8], [9], [10], [11], [12], [13] y [14]. Como resultado se elaboran documentos que sirven para el desarrollo del sistema, conocer el estado del arte en el tema y las alternativas existentes como el producto OpenEmpi[15]. Se sigue un proce-

so continuo, iterativo e incremental que pasa por varias etapas de refinamiento y validación. Como resultado de este proceso se obtiene la versión del Índice Maestro de Pacientes que se utiliza actualmente en producción en el Hospital Regional Antonio J. Scaravelli. El proceso iterativo utilizado para gestionar y concretar el proyecto se basa en Scrum[16].

El sistema presentado en este trabajo se basa en el estudio de las características de [1] y [3]. Este trabajo se diferencia de estas propuestas en los componentes Java EE[7], desarrollados para el módulo búsqueda de candidatos. Antes de comenzar el proyecto fue evaluado el software openempi, se analizó parte de su diseño, pero se decantó por uno propio por las siguientes razones: escaso soporte de Auditoría e interfaz de usuario con funcionalidad muy limitada.

2.2 Arquitectura del Sistema

En base a los objetivos planteados antes se presenta una propuesta para la implementación de un Índice Maestro de Pacientes, utilizando una arquitectura compuesta por software open source. Presentando como principal aporte el diseño y la implementación realizada para el desarrollo del módulo de búsqueda de candidatos del Índice maestro de Pacientes.

La tecnología utilizada se centra en Java EE utilizando como componente principal el software Hibernate Search[8] con apache Lucene[9]. En la Figura 1, se muestra el esquema con la tecnología utilizada. Se utiliza hibernate search como motor de búsqueda en texto completo por las siguientes razones: Alta integración con el api de persistencia de java, rápida indexación en disco, soporte nativo en servidor de aplicaciones JBOSS y gran variedad de algoritmos de bloqueo y similitud. El software desarrollado está construido basado en la arquitectura MVC (Modelo Vista Controlador). Los framework más destacables en cada capa son los que se muestran en la siguiente figura:

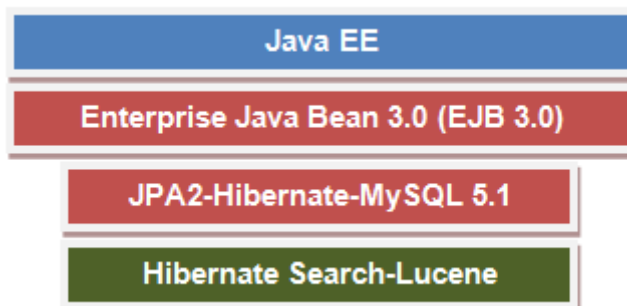


Fig. 1. Frameworks utilizados

2.3 Características del Recurso Paciente

Para identificar a una persona hay que usar atributos que perduren o cambien poco en el tiempo. En este trabajo se utilizan algunos de los propuestos en [3], los cuales están formados por: apellido, nombre, sexo, fecha de nacimiento, tipo y número de documento de identidad. A los atributos que se utilizan para la identificación única de una persona se denomina conjunto mínimo de datos. El uso de un conjunto básico de atributos es conveniente complementarlo con el uso de algún tipo de registro biométrico como la foto de la cara o más avanzados como el de retina o reconocimiento de rostro. El uso de registros biométricos permite la identificación unívoca de pacientes en forma rápida y precisa. En la Figura 2, se puede observar un ejemplo de conjunto mínimo de datos en la pantalla de búsqueda implementada en el sistema.

Fig. 2. Ejemplo de conjunto mínimo de datos

Se definen los estados activo e inactivo para los objetos pacientes. El activo es aquel paciente que está en condiciones de ser atendido. Este tipo de pacientes puede estar en alguno de los siguientes estados:

Permanente. Son aquellos pacientes que:

- Están ingresados al padrón aportando el conjunto de información denominada “Datos Permanentes”.
- Son datos que se considera inalterable a lo largo de la vida del individuo.
- Su combinación es la base para la identificación unívoca del sujeto.

Validado. Se definen así a aquellos pacientes:

- Con quienes un Auditor ha logrado un contacto personal o telefónico ulterior al contacto personal con el Empadronador, confirmando el conjunto de “Datos Permanentes”.
- Son datos de máxima confiabilidad.
- Se recaba información adicional que es útil a los fines asistenciales en la Institución, pero que son variables a lo largo del tiempo o pueden estar ausentes.

Temporario. Es el estado que toman los pacientes que:

- No se cuenta con los “Datos Permanentes” y requieren atención médica sin demora.
- Es necesario realizar al menos una búsqueda entre los Pacientes que ya se encuentran en el Padrón.

Histórico: Se utiliza cuando comienza la implementación de un Índice Maestro de Pacientes. Los registros que se toman de las base de datos de personas son cargados inicialmente como objetos o registros históricos.

Los Pacientes **Inactivos** no se tienen en cuenta en las búsquedas. Un paciente se encuentra en estado Inactivo cuando:

- Siendo un paciente Temporario ha sido Rechazado por el Auditor.
- El fallecimiento de un paciente Activo. Al colocarle fecha de fallecimiento se modifica a Paciente Inactivo automáticamente.
- Por Fusión de pacientes Activos.

2.4 Componentes Fundamentales de la Aplicación

Los componentes principales del Índice Maestro de Pacientes están formados por los siguientes módulos.

Indexación de Registros.

Para realizar este proceso se integra a nuestra aplicación el motor de búsquedas Hibernate Search, que permite indexar el contenido de nuestro modelo de datos para ofrecer unas búsquedas avanzadas con texto libre. Se utiliza como analizador fonético el algoritmo Double Methafone[12] configurado para su uso con el idioma español. Este método permite por cada cadena, obtener un código que permite su normalización para su posterior comparación con otras cadenas. El índice se arma utilizando el algoritmo Double Methafone sobre la cadena formada por la unión de los atributos Apellido y Nombre.

Bloqueo (Bloquing)

La implementación de un MPI tiene varios desafíos; uno de los principales es la búsqueda de un conjunto mínimo de datos sobre grandes cantidades de datos. En una base de datos con 'n' registros el problema se plantea como de complejidad $O(n^2)$ [10]. El método de fuerza bruta es el más fiable pero el menos efectivos debido a que la comparación de registro por registro requiere de $(n^2-n) / 2$ comparaciones.

En este trabajo se utiliza una implementación de una técnica de bloqueo basado en un criterio que utiliza como variable el apellido y nombre del paciente. Entonces se busca en el índice armado en el punto anterior por la cadena apellido+nombre del paciente obteniendo un subconjunto de 200 candidatos con gran velocidad.

Cálculo de Pesos

Teniendo en cuenta los candidatos obtenidos en el punto anterior se utiliza el algoritmo Jaro-Winkler para medir el grado de similitud entre dos cadenas. La puntuación que devuelve el algoritmo se normaliza de forma que 0 equivale a ninguna similitud y 1 es una coincidencia exacta. Se aplica por separado el algoritmo a cada uno de los elementos del conjunto mínimo de datos y luego se multiplica por un peso que pondera a los atributos. La suma de lo anterior es el peso resultante que es una medida de la similitud entre el conjunto mínimo de datos de un paciente y el conjunto mínimo de datos de cada uno de sus objetos candidatos.

3 Utilización en Múltiples Organismos de Salud

Para cumplir con el objetivo de soportar la identificación de pacientes a través de múltiples organismos de salud se implementa el esquema propuesto por CORBAMED PIDS [17]. Cada organismo de salud asigna identificadores (ID) que identifican a pacientes dentro de su dominio local de valores de ID. Fuera de ese sistema u organización dichos IDs carecen de significado. Mediante la implementación realizada se puede soportar de manera simultánea la asignación de IDs dentro de un dominio particular y la correlación de IDs entre múltiples dominios. En la figura 3, se muestra un ejemplo representativo de la implementación realizada. El paciente Juan tiene un identificado único correlacionado con los ID del resto de las instituciones de salud por donde ha recibido atención médica.

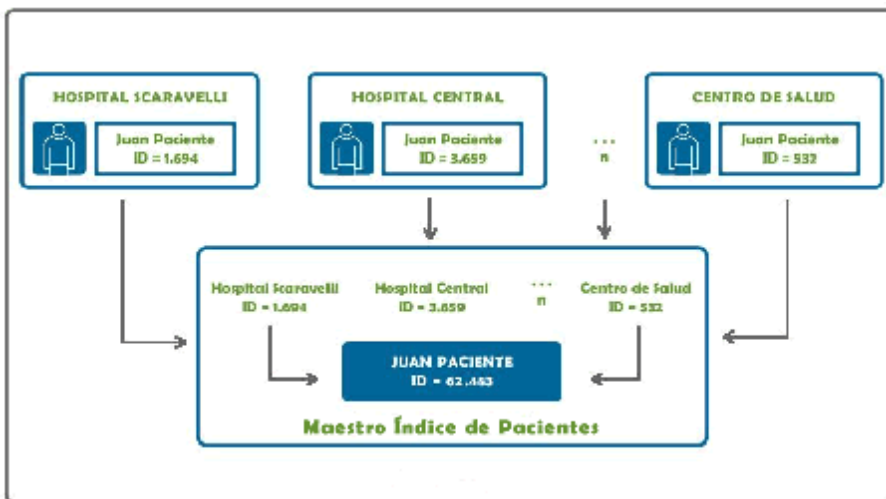


Fig. 3. Identificación de Pacientes en Múltiples Organismos

4 Roles y Procedimientos Implementados

Los roles con los que trabaja el sistema son los de empadronador y auditor. Cada uno de los cuales realiza las tareas que se detallan a continuación:

4.1 Empadronador

Recibir al paciente y acreditar su identidad. Se solicita algún documento de identidad a la persona para validar su identidad. En la Figura 4, se observan los elementos involucrados.



Fig. 4. Acreditación de Identidad con Documento

- Identificar si el Paciente no se encuentra previamente incorporado al Padrón. Se explica en el punto
- Incorporar al Padrón de pacientes. En la Figura 5, se muestra un ejemplo.
- Actualización y/o corrección de datos registrados en el Padrón.



Fig. 5. Alta y/o Actualización de Datos Pacientes en el Padrón

4.2 Auditor

- Autenticar Pacientes Temporarios, Validar Paciente, Inactivar Paciente, Reactivar Paciente, Rechazar Paciente, Búsqueda de posibles duplicados. En la Figura 6, se observa un ejemplo.

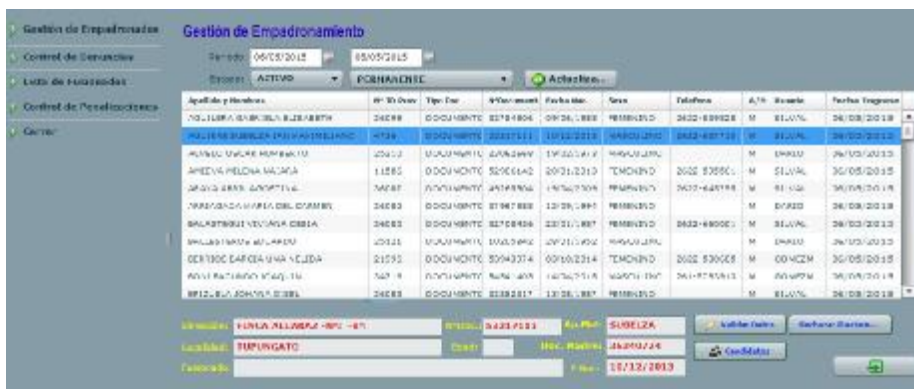


Fig. 6. Gestión de Empadronamiento

- Fusionar Paciente. La fusión de uno o más pacientes consiste en la unificación de sus registros en un solo paciente validado. En la Figura 7, se muestran dos objetos duplicados a los que se les aplica el proceso de fusión.



Fig. 7. Fusión de Pacientes

- Revisión de los pacientes Temporarios nuevos creados.
- Control de calidad de la información del Padrón (números de teléfono inválidos, nombres inválidos, error en la asignación del sexo o la fecha de nacimiento).
- Control de denuncias de los Empadronadores. Los empadronadores pueden realizar denuncias si detectan posibles duplicados en la búsqueda de candidatos. El auditor revisa las denuncias en base a la pantalla que se muestra en la Figura 8.

Apellido y Nombre	N°HC	Fecha Nac	Sexo	Tipo Dec	Documento	Usuario denuncia
ACEVEDO BLANCO CELINA VIOLETA	18613	19/01/2000	FEMEF	DOCUMENTO U	42750834	MARADONA
ACEVEDO CELINA	30182	19/01/2000	FEMEF	INDOCUMENTA	30182	MARADONA
BELARDEZ SILVANA PATRICIA	25583	05/08/1997	FEMEF	DOCUMENTO U	42751224	ANTONIO
BENITES MARIA BELEN	14281	22/09/1994	FEMEF	DOCUMENTO U	38378997	ANTONIO
BENITEZ MARIA BELEN	2250	22/09/1994	FEMEF	INDOCUMENTA	2250	ANTONIO
CANZAMPI RARI	NAN	02/04/2013	MASC	EN TRAMITE (R)	0	MBARRERA
CAIGURA ADRIANA ANDREA	NAN	23/10/1966	FEMEF	DOCUMENTO U	93021083	MARADONA
CASHJARA MAMANI ADRIANA ANOMBA	25754	23/10/1966	FEMEF	DOCUMENTO U	93021085	MARADONA
CAMPOS GRACIELA ROSANA	29334	10/10/1964	FEMEF	DOCUMENTO U	17128327	ANTONIO
CAMPOS GRACIELA ROSAURA	17528	10/10/1964	FEMEF	DOCUMENTO U	27128327	ANTONIO

Fig. 8. Control de Denuncias

5 Proceso de Búsqueda de Pacientes

En base a los tres componentes de software desarrollado se puede implementar el proceso de búsqueda de candidatos que es esencial para la detección de datos duplicados. Aquellas instancias que se detectan como duplicados son denunciadas para su posterior verificación por auditor.

En la figura 9, se muestra un ejemplo de búsqueda. En la parte de arriba de la pantalla se coloca el conjunto mínimo de datos de la persona que se quiere buscar para realizar operación de consulta, modificación o alta de pacientes. Abajo, el sistema devuelve el conjunto de candidatos con sus pesos respectivos.

Como se puede apreciar en la figura las dos primeras instancias de candidatos tienen el mismo peso 0,95. Esto indica que existen altas probabilidades de que los datos del paciente estén duplicados. En este caso el usuario empadronador del sistema tiene que seleccionar ambas instancias similares y denunciarlos para que sea revisado luego por el auditor.

Apellido y Nombre	ID Pro	Edad	Fecha Nac	Sexo	Tipo Doc	Documento	Peso	Dominio	Id Interno
CABRERA LORENA ELIZABETH	0	28	03/01/1985	FEMEN	DOCUME	31994191	0,950000	6-HS Antonio J. Sci	66051
CABRERA LORENA ELIZABETH	0	28	03/01/1985	FEMEN	DOCUME	31491194	0,950000		
CABRERA ANA	0	60	03/01/1953	FEMEN	DOCUME	10573358	0,897828		
CABRERA CORDOBA MARIA ELIZABETH	0	31	03/02/1982	FEMEN	DOCUME	29340968	0,871193		

Fig. 9. Pantalla de Búsqueda

6 Conclusiones y Trabajo Futuro

En base al problema de duplicación de registros de pacientes planteado en los organismos de salud de la provincia de Mendoza. Se investigan distintas opciones y se decide el desarrollo de un Índice Maestro de Pacientes que se implementa como prototipo en él. El prototipo evoluciona posteriormente a un sistema en producción y en la actualidad, se está utilizando como base para el desarrollo del Índice Maestro de Pacientes de la provincia de Mendoza.

El sistema funciona en el Hospital Regional Antonio J. Scaravelli, desde octubre del año 2013. La base de datos de pacientes cuenta con 140.000 registros. Desde esa fecha se han realizado 2429 denuncias. En base a las denuncias se han corregido los datos de 2346 registros y se han fusionado 1081 registros.

El desarrollo realizado está basado en estándares internacionales de salud y permite cumplir con los requerimientos planteados. Esto es principalmente lograr disminuir la duplicación de registros en instituciones de salud y unificar los datos mediante un servicio de identificación de personas. El módulo de auditoría de datos implementado es esencial para el funcionamiento del sistema. El aporte principal del trabajo presentado en este documento, está en el módulo de búsqueda de candidatos donde se implementan algoritmos que funcionan utilizando como base las consultas de objetos indexados.

Referencias

1. Garfi L., Navajas P., Gómez A., Luna D., González Bernardo de Quirós F: Implementación de un sistema centralizado para la identificación de pacientes en un hospital de alta complejidad. 5to Simposio de Informática en Salud - 31 JAIIO. – 2002.
2. Appavu, S., Analysis of Unique Patient Identifier Options. 1997, The National Committee on Vital and Health Statistics (NCVHS) - Department of Health and Human Services (U.S.).
3. Weber, G.I. Achieving a patient unit record within electronic record systems. in Toward an Electronic Patient Record. 1995: Newton, Mass.
4. Hospital Regional Antonio J. Scaravelli. www.hospital-scaravelli.mendoza.gov.ar. Último acceso 20/04/2015.
5. C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell and A. Wesslén, "Experimentation in Software Engineering", Springer, ISBN 978-3-642-29043-5, 2012.
6. Marcos, E., "Investigación en Ingeniería del Software vs. Desarrollo", Grupo KYBELE. Universidad Rey Juan Carlos. <http://gidis.inf.pucp.edu.pe/recursos/InvIngSWvsDS.pdf>
7. JSR-000316 Java Platform, Enterprise Edition 6 Specification. Version 6.0. SUN MICROSYSTEMS, INC. 2009.
8. Hibernate Search. www.hibernate.org/search. Último acceso 20/02/2015.
9. Apache Lucene. www.lucene.apache.org/core. Último acceso 20/02/2015.
10. Erik A Sauleau, Jean-Philippe Paumier, Antoine Buemi. Medical record linkage in health information systems by approximate string matching and clustering. 2005 Sauleau et al licensee BioMed Central Ltd. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1274322/>
11. The OMG CORBAed Domain Task Force, Person identification service (PIDS). 1998.
12. L. Philips, "Hanging on the metaphor," Computer Language, vol. 7, no. 12 (December), 1990.

13. IVAN AMON URIBE, "Funciones de similitud sobre Cadenas de Texto: Una comparación basada en la Naturaleza de los datos." En: Jamaica. 2010. Evento: International Conference on Information Resources Managment (CONF-IRM2010). Montego Bay, Jamaica. Ponencia: Funciones de similitud sobre Cadenas de Texto: Una comparación basada en la Naturaleza de los datos. Libro:, , p. - , v. <, fasc.
14. IVAN AMON URIBE, "Detección de Duplicados: Una Guía Metodológica." En: Colombia. 2010. Evento: Quinto Congreso Colombiano de Computación 5CCC Ponencia: Detección de Duplicados: Una Guía Metodológica. Libro:, , p. - , v. <, fasc.
15. OpenEmpi. www.openempi.org. Último acceso 15/02/2015.
16. Scrum. www.scrum.org. Último acceso 15/02/2015.
17. Person Identification Service (PIDS) Specification. www.omg.org/spec/PIDS/1.1. Último acceso 10/10/2014.